

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 16:14:38 ; Search time 7193 Seconds
(without alignments)
10965.350 Million cell updates/sec

Title: US-09-890-811B-9
Perfect score: 1928
Sequence: 1 gcaaggtctcgtgcattct.....tcaaaaaaaaaaaaaaaa 1928

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.in.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pin.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1928	100.0	1928	6	AX192162	AX192162 Sequence
2	549.4	28.5	1993	8	IPBSPP1P	D30038 Ipomoea bat
3	489.6	25.4	78249	8	AP004529	AP004529 Lotus jap
4	454.2	23.6	1823	8	AB020590	AB020590 Nicotiana
5	430.6	22.3	1988	8	AB063573	AB063573 Nicotiana
6	424.6	22.0	1867	8	AB022693	AB022693 Nicotiana
7	388.6	20.2	1809	8	AF080595	AF080595 Pimpinell
8	379	19.7	1800	8	PCU48831	U48831 Petroselinu
9	353.4	18.3	1417	8	AB063575	AB063575 Nicotiana
10	308.2	16.0	1539	6	AX412575	AX412575 Sequence
11	308.2	16.0	1539	6	AX412905	AX412905 Sequence
12	308.2	16.0	1539	6	AX506109	AX506109 Sequence
13	308.2	16.0	1539	6	AX651760	AX651760 Sequence
14	308.2	16.0	1539	8	AF509499	AF509499 Arabidops
15	277.2	14.4	1668	6	AX654743	AX654743 Sequence
16	270.6	14.0	1213	8	BT002338	BT002338 Arabidops
17	270.6	14.0	1264	8	AY114650	AY114650 Arabidops
18	270.6	14.0	1453	8	AY062720	AY062720 Arabidops
19	270.6	14.0	1462	8	AY045813	AY045813 Arabidops
20	270.6	14.0	1606	8	AY1136318	AY1136318 Arabidops
21	270.2	14.0	1182	6	AX506986	AX506986 Sequence
22	270.2	14.0	1182	6	AX652016	AX652016 Sequence
23	270.2	14.0	1182	8	AF418309	AF418309 Arabidops
24	270.2	14.0	1182	8	BT008482	BT008482 Arabidops
25	237.8	12.3	1785	8	AX653781	AX653781 Sequence
26	234.6	12.2	1459	8	AFABF1	Z48429 A. fatua mRN
27	228.8	11.9	1542	6	AX505730	AX505730 Sequence
28	228.8	11.9	1542	6	AX589855	AX589855 Sequence
29	228.8	11.9	1895	8	AF224701	AF224701 Arabidops
30	221.6	11.5	2158	6	AX192164	AX192164 Sequence
31	221.4	11.5	1489	8	AY150436	AY150436 Arabidops
32	221.4	11.5	1545	6	AX507584	AX507584 Sequence
33	221.4	11.5	1545	6	AX589831	AX589831 Sequence
34	221.4	11.5	1545	8	AF425835	AF425835 Arabidops
35	221.4	11.5	1545	8	BT002629	BT002629 Arabidops
36	221.4	11.5	1736	8	AY045892	AY045892 Arabidops
37	221.4	11.5	1837	8	AY045676	AY045676 Arabidops
38	221.4	11.5	1872	8	AF425837	AF425837 Arabidops
39	221.4	11.5	1969	8	AF224703	AF224703 Arabidops
40	208.2	10.8	1626	6	AX463546	AX463546 Sequence
41	207.8	10.8	2160	8	AF096299	AF096299 Nicotiana
42	206.4	10.7	2073	8	AF140554	AF140554 Avena sat
43	204	10.6	1719	6	AX654232	AX654232 Sequence
44	187.8	9.7	1787	8	AF439274	AF439274 Retama ra
45	187.6	9.7	2086	6	AX192160	AX192160 Sequence

ALIGNMENTS

RESULT 1	AX192162	AX192162	1928 bp	DNA	linear	PAT 15-AUG-2001
LOCUS	Sequence 9 from Patent WO0149840.					
DEFINITION	AX192162					
ACCESSION	AX192162					
VERSION	AX192162.1	GI:15210203				
KEYWORDS	Glycine max (soybean)					
SOURCE	Glycine max					
ORGANISM	Glycine max					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
	rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;					
	Glycine.					
REFERENCE	1					
AUTHORS	Zhu,Q., Famodu,O.O. and Rafalski,J.A.					

TITLE Spfl-related transcription factors
JOURNAL Patent: WO 0149840-A 9 12-JUL-2001;
E.I. DU PONT DE NEMOURS AND COMPANY (US)
FEATURES Location/Qualifiers
source 1..1928
/organism="Glycine max"
/mol_type="genomic DNA"
/db_xref="taxon:3847"
BASE COUNT 591 a 454 c 387 g 496 t
ORIGIN

Query Match 100.0%; Score 1928; DB 6; Length 1928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCACGAGTCTCAGGCAATCTTCTTCTGTTAGTTAGACACCTCTGCAAGTGCAGAACTCCT 60
Db 1 GCACGAGTCTCATGGCAATCTTCTTCTGTTAGTTAGACACCTCTGCAAGTGCAGAACTCCT 60

Qy 61 TCACCAACTTCACTTCTCCACACACCTTTCATGACCACTTCTTCTGACCTCTCTG 120
Db 61 TCACCAACTTCACTTCTCCACACACCTTTCATGACCACTTCTTCTGACCTCTCTG 120

Qy 121 CTTCTCCCTTGGCAACAAACAGCCACACAGGCTGTTTCTCTGAGAGAACTGGCTCTG 180
Db 121 CTTCTCCCTTGGCAACAAACAGCCACCA CAGGGTGGTTTGTCTGAGAGAACTGGCTCTG 180

Qy 181 GTGTTCCCAAAATCAAGTCCACACCACTTCTGCTCTCTCTCTCCCTCCCAATTT 240
Db 181 GTGTTCCCAAAATCAAGTCCACACCACTTCTGCTCTCTCTCTCCCTCCCAATTT 240

Qy 241 CTCCTTCTTCTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 300
Db 241 CTCCTTCTTCTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 300

Qy 301 CGGTTCTCTTAACTCTTCTTCAACAACTTCTGCCATCTCCAACTCTGAGAGCAATTTGTTGCTC 360
Db 301 CGGTTCTCTTAACTCTTCTTCAACAACTTCTGCCATCTCCAACTCTGAGAGCAATTTGTTGCTC 360

Qy 361 AGAGCTTCAATTTGGAAGAGCACTTCTGAGGGAATCAGCAAAATTTGTCAAGGAAGAGCA 420
Db 361 AGAGCTTCAATTTGGAAGAGCACTTCTGAGGGAATCAGCAAAATTTGTCAAGGAAGAGCA 420

Qy 421 AAAGCTTCTCAAAATTTCTTCTTCAAACTCCGATCAGGACCTCTCTCTCTCTCTCTCTCT 480
Db 421 AAAGCTTCTCAAAATTTCTTCTTCAAACTCCGATCAGGACCTCTCTCTCTCTCTCTCTCT 480

Qy 481 CATACCACTTCTCAAAATTTCTCAAGTTCACAGTTCACACACACAGCCATGGAGTTTTCAGGAGGCCA 540
Db 481 CATACCACTTCTCAAAATTTCTCAAGTTCACAGTTCACACACACAGCCATGGAGTTTTCAGGAGGCCA 540

Qy 541 CGAAACAGGATAATTTTCTCTCAGGAAGGGTATGATGAAACCTGAAAACTCTTCTTCTCA 600
Db 541 CGAAACAGGATAATTTTCTCTCAGGAAGGGTATGATGAAACCTGAAAACTCTTCTTCTCA 600

Qy 601 TGCAGAGTTTCTCCCTGAGATTCTGATGTCGAACTCAAACTAACCATAGCAATGGGTTTCAAT 660
Db 601 TGCAGAGTTTCTCCCTGAGATTCTGATGTCGAACTCAAACTAACCATAGCAATGGGTTTCAAT 660

Qy 661 CCGATTATGGCAATTTACCCCTCAATCTCAGACTTTTAAAGTAGAAGGTTCAGATGATGGGT 720
Db 661 CCGATTATGGCAATTTACCCCTCAATCTCAGACTTTTAAAGTAGAAGGTTCAGATGATGGGT 720

Qy 721 ACAATTGGAGAAATATGGCCCAAAACAAAGTGAAGGGAAGTGAATAATCCAAAGATTATT 780
Db 721 ACAATTGGAGAAATATGGCCCAAAACAAAGTGAAGGGAAGTGAATAATCCAAAGATTATT 780

Qy 781 ACAATGCAATACCCCAATTTGCCCTCAAGAGAGAGAGGTTCAGAGCTCTTTAGATGGAC 840
Db 781 ACAATGCAATACCCCAATTTGCCCTCAAGAGAGAGAGGTTCAGAGCTCTTTAGATGGAC 840

Qy 841 AAATTTACTGATAGTTTATAGGGTACTCATAACTCCTTAAGCTCTCAAAATCTAGGA 900
Db 841 AAATTTACTGATAGTTTATAGGGTACTCATAACTCCTTAAGCTCTCAAAATCTAGGA 900

841 AAATTTACTGATAGTTTATAGGGTACTCATAACTCCTTAAGCTCTCAAAATCTAGGA 900
901 GAAACTCATCAAACTCCTCTCTCTTGCATCCCTCATTAATCAATTCATCAGAACTGAAA 960
901 GAAACTCATCAAACTCCTCTCTCTTGCATCCCTCATTAATCAATTCATCAGAACTGAAA 960
961 TCCAGATCAATCCTATGTCACACATGGAAGTGGACAAATGGATTCAGCTGCCACCCAG 1020
961 TCCAGATCAATCCTATGTCACACATGGAAGTGGACAAATGGATTCAGCTGCCACCCAG 1020

1021 AAAACTCATCAATATCAATTTGAGATGATGATTTTGGAGAGAGTTTCCAAAAGTGTAAAT 1080
1021 AAAACTCATCAATATCAATTTGAGATGATGATTTTGGAGAGAGTTTCCAAAAGTGTAAAT 1080

1081 CAGGAGGGATGAATATGATGAAGATGAACCTGTATGCCAAAAGATGGAATAATGAAGGTG 1140
1081 CAGGAGGGATGAATATGATGAAGATGAACCTGTATGCCAAAAGATGGAATAATGAAGGTG 1140

1141 AAAATGAGGGTATGTGAGCCCTCTGGAAGTAGAACCTGAGAGAACCTAGAGTTGTAGTTT 1200
1141 AAAATGAGGGTATGTGAGCCCTCTGGAAGTAGAACCTGAGAGAACCTAGAGTTGTAGTTT 1200

1201 AGACAAACAGTGAATGATATCTTCTGATGATGGCTATAGGTGGAGAAAATACGGGCAGA 1260
1201 AGACAAACAGTGAATGATATCTTCTGATGATGGCTATAGGTGGAGAAAATACGGGCAGA 1260

1261 AGTAGTGAAGGCAATCCAAATCCAAAGGATTTACTACAAAGTGCACACACCCAGGATGTC 1320
1261 AGTAGTGAAGGCAATCCAAATCCAAAGGATTTACTACAAAGTGCACACACCCAGGATGTC 1320

1321 CAGTGAAGGAAGCAGCTGGAAAGAGCCTCACATGACCTAAGGGCTGTGATCACAACCTTATG 1380
1321 CAGTGAAGGAAGCAGCTGGAAAGAGCCTCACATGACCTAAGGGCTGTGATCACAACCTTATG 1380

1381 AGGAAAGACACAACTATGATTTCTGAGCCCTGGAGTGGAGAGCAATCTCTGTAACA 1440
1381 AGGAAAGACACAACTATGATTTCTGAGCCCTGGAGTGGAGAGCAATCTCTGTAACA 1440

1441 GACCAATGCCAAACAAATGCTTCAACACCAACCACTGAGCCACTTCCGTAAGGCTCT 1500
1441 GACCAATGCCAAACAAATGCTTCAACACCAACCACTGAGCCACTTCCGTAAGGCTCT 1500

1501 TGCCAGTGTATCCACCAAGTGAACAATTCCTTTCAGAACCAAGATCACAAGCACCACCG 1560
1501 TGCCAGTGTATCCACCAAGTGAACAATTCCTTTCAGAACCAAGATCACAAGCACCACCG 1560

1561 AAGGCAATCACCTTTCAGCTAGAGATGCTA CAAAAGTCCAGGAAGTTTGGATTTCTCAG 1620
1561 AAGGCAATCACCTTTCAGCTAGAGATGCTA CAAAAGTCCAGGAAGTTTGGATTTCTCAG 1620

1621 GGTTCGGATCCAATGCAATCTTACGTGAACCAAGCACTATCTGCAATGTTTCT 1680
1621 GGTTCGGATCCAATGCAATCTTACGTGAACCAAGCACTATCTGCAATGTTTCT 1680

1681 CTTCCAGGACCAAGGAGGAGCCTTAGAGATGACATGTTTCTTGTAGTCTTACTATGCTGAA 1740
1681 CTTCCAGGACCAAGGAGGAGCCTTAGAGATGACATGTTTCTTGTAGTCTTACTATGCTGAA 1740

1741 GGAATTTTTTTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1800
1741 GGAATTTTTTTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1800

1801 CTAGGATTTATTCGACAAATAGGTTCCATTCGATTTATGCAATTTTCTGTTTCTTCTTCT 1860
1801 CTAGGATTTATTCGACAAATAGGTTCCATTCGATTTATGCAATTTTCTGTTTCTTCTTCT 1860

1861 TTGTAAATTTTATACAGCCACAGGATTTGGTATAGTATATATAGTATTTTCAAAAAAAA 1920
1861 TTGTAAATTTTATACAGCCACAGGATTTGGTATAGTATATATAGTATTTTCAAAAAAAA 1920

1921 AAAAAAA 1928
1921 AAAAAAA 1928

RESULT 2
 LOCUS IPBSPF1P 1993 bp mRNA linear PLN 29-JAN-2003
 DEFINITION Ipomoea batatas mRNA for SPF1 protein, complete cds.
 ACCESSION D30038
 VERSION D30038.1 GI:484260
 KEYWORDS

SOURCE Ipomoea batatas (sweet potato)
 ORGANISM Ipomoea batatas
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; lamids; Solanales; Convolvulaceae; Ipomoea.

REFERENCE 1 (bases 1 to 1993)
 AUTHORS Ishiguro, S. and Nakamura, K.
 TITLE Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and beta-amylase from sweet potato
 JOURNAL Mol. Gen. Genet. 244 (6), 563-571 (1994)

PUBMED 95058910
 REFERENCE 2 (bases 1 to 1993)
 AUTHORS Ishiguro, S.
 TITLE Direct Submission
 JOURNAL Submitted (26-APR-1994) Sumie Ishiguro, National Institute for Basic Biology, Div. 1 of Gene Expression & Regulation; 38 Nishigonaka, Myodaiji-cho, Okazaki, Aichi 444, Japan (E-mail: guroyan@nibb.ac.jp, Tel: 81-564-55-7682, Fax: 81-564-55-7685)

FEATURES
 source
 1. 1993
 /organism="Ipomoea batatas"
 /mol_type="mRNA"
 /cultivar="Kokei No.14"
 /db_xref="taxon:4120"
 /clone="SPF1"
 /tissue_types="petiole"
 /note="common name: sweet potato"
 29..1678
 /codon_start=1
 /product="SPF1 protein"
 /protein_id="BAA06278.1"
 /db_xref="GI:484261"
 /translation="MAASGTIDAPTASSFSFSTASSFMSPTDLLASDAYSGSVS
 RGLGRIARTGSGVPKFKSLPLPSLPLASPAVSPSYFAFPPLGLSPSLDPSVLLS
 SSNLSPTGTGTPAQTFWKNSNASQEDVKQERGYDFDFQTSNSMTLNYEDSK
 RKDELNSQLPVTSTQMSQNGGSSEYNNCCPPSTLREQRSDDDGYNWRKY
 GQKVQSENPRSYKTHPCNPKKKKVERALDQITEIVYGAHNHPKPSRTRSSS
 STASSASTLAAQSYNAPASDPDQSYMSNGQMSVATPENSISGVDFEFSQSK
 RESGDFDEPDPAKRWKVENEGVSAQGSRTVREPVVQTTSDIDILDDGYWR
 KYGQKVGNPNRPSYKCTSGQCPVKHVERASHDIRSVITTYEGKHNDVPAEGRS
 GSGLNRGANPNNNAMAMAIRSTMSLQSNYPIPISTRPMQGGQAPYEMLOGSG
 GFGYSGFGNPMNAYANOIQDNFASRAKEPRDRLDFTLLA"
 656..775
 /note="direct repeat"
 1184..1303
 /note="direct repeat"

repeat_region 471 c 495 g 516 t

repeat_region 471 c 495 g 516 t

polya_site 471 c 495 g 516 t

BASE COUNT 511 a 471 c 495 g 516 t

ORIGIN

Query Match 28.5%; Score 549.4; DB 8; Length 1993;

Best Local Similarity 61.2%; Pred. No. 2.3e-122;

Matches 1070; Conservative 0; Mismatches 571; Indels 108; Gaps 7;

QY 8 TCTCATGGCATCTTCTTCTGGTGTAGTTAGACACCTCTGCAAGTGCACAACTCTTCCACCA 67

DB 25 TCTAATGGCTGCTCTTTCAGGACATAGAGCCGCCACACAGCTTCTCATCTTCTCTTT 84

QY 68 CTTACCTTCTCCACACACCCCTTTCATGACACACTTCTTCTGACCTCTGCTCTCTCC 127

DB 85 CTCACCGCTCTTCTCATGTCTCTTCTTCACTGACCTCTCTGCTCTCCGACGCTATTC 144

QY 128 CTTTGGACAACAACAAGCCACCAAGGGTG---GTTTGTCTGAGAGAACTGGCTCTGGTGT 184
 DB 145 CGCGGCTCTGTGAGCAGAGGGCTTGGTGTATCGGATAGCGGAGAGAACTGGTCTGGTGT 204
 QY 185 TCCCAAAATTCAGTCCACAGCACCACCTTCTCTGCTCTCTCTCTCCCTCCCAATTTCTCC 244
 DB 205 GCCCAAGTTTAACTCTTTTGGCGCGCGCTCTCTGCGCGCTTTCTGCGCGCGCTCTCGCC 264
 QY 245 TTCTTCTTACTTTTGTCTATTCCTCTCTGAGCCCTGCTGAGCTTCTTGAATCTGCGGT 304
 DB 265 GTGCTCTTACTTCTGCTTTTCCGCTGGTGTAGCCCGCAGTGTCTCTGATTTCCCTGT 324
 QY 305 TCTCTTAACTCTTCTCAACAATTTCTGCAATCTCTCAACAACCTGGAGCAATTTTGTCTCAGAG 364
 DB 325 TCTTCTATCTCTCTCAACAATTTTGGCGCTCTCAACTACTTGGGACTTTTCTCTCTCAGAC 384
 QY 365 CTTCAATTCGAAGAGCAGTTTTCAGGGGGGAATCAGCAAAATTTCAAGAGAGAGACAAAAG 424
 DB 385 CTTCAACTGGAAGAATGATTTCTAAACGCATCCCAAGGAAGATGTTAAGCAAGAGAGAAAG 444
 QY 425 CTTTCTCAAAATTTCTCTTCCAAACCGGATCAGGACCTCTGCTTCTCATCCACAGCAACATA 484
 DB 445 ATACCAGATTTCTCTTTTCAGACCACT-----CTGCTTCAATGACATGTAATTA 495
 QY 485 CCAGTCTTCAAAATGTCTCAGTTTCAACAACAACAGCCATCGAGTGTTCAGGAGGCCACGAA 544
 DB 496 TGAAGATTCAAAGAGAGAAAGATGAGCTCAATTTCTCTGAGAGGCTTTCCCTCTGTA--- 551
 QY 545 ACAGGATATTTTCTCTCAGGAAAGGATGATGATAAACTGTAAACTCTTCTTCTCCATGCA 604
 DB 552 -----CTACTTCAACTCA 564
 QY 605 GAGTTTTCCTCCCTGAGTGTGTCCAAATAACCATAGCAATGGGTTTCAATCCGA 664
 DB 565 GATGAGCTCTCAGACAATGGTGGGAGTACTCTGAGTATAATA-----A 609
 QY 665 TTATGCAATTAACCCCTCACAATCTCAGACTTTTAACTAGAGGTGAGATGATGGGTACAA 724
 DB 610 TCAATCTCGCCCTTCTCCAGAGCTTTGAGGGAGCAGAGCGATCTGATGACGGGTACAA 669
 QY 725 TTGAGGAAATATGGCCAAAAACAAGTGAAGGAAGTGAATAATCCAAGAGTTATTATCAA 784
 DB 670 TTGAGGAAATATCGGCGAGAAACAGGTGAAGGGAGCGAAAAATCCGAGGAGTTATTATCAA 729
 QY 785 ATGCACATACCCCAATTTGCCCTACAAAGAGAGGTGAGAGTCTTTAGATCGACAAAT 844
 DB 730 GTGTACGACCCCAATTTGCCCAAGAAAGAGGTGAGAGGGCTTTTGGATGGGAGAT 789
 QY 845 TACTGAGATAGTTTAAAGGTACTCATAAACCATCTTAAGCCCTCAAAATACTAGGAGAAA 904
 DB 790 TACTGAGATGTCTACAAGAGAGCTCACAAATCACCCGAGAGCTCAGTCCACTAGGAGATC 849
 QY 905 CTCATCAAACTCTCTTCT-----CTTGAATCCCTCATTTCAAAATTCATCAG 952
 DB 850 GTGCTCTCTACAGCTTCTTTCGCTTCAACCTTGGCTGCCAGTCTTATAATTCGCTGC 909
 QY 953 AACTGAAATCCAGATCAATCTATGCCACATGGAAGTGGACAAATGATTCAGTCTC 1012
 DB 910 CAGTGATGTCGGGATCAGTCTGCTCTAATGTGAACGGGAGATGATTTCTGTTC 969
 QY 1013 CACCCAGAAAACTCATCAATATCAATTTGAGATGATGATTTTGGACAGAGTTTCCCAAAA 1072
 DB 970 CAGCCAGAGATTTCTTCGATCTCGTGGGGATGATGATTCGACAGAGCTCTCAGAA 1029
 QY 1073 GTGTAATCAGAGGGGATGAATATGATGAAGATGAACCTGATGCCAAAGATGGAAGAT 1132
 DB 1030 GAGGAGTCCGGGGAGACGAGTTTGTGAAGACGAAACCGGATGCAAGAGATGGAAGAT 1089
 QY 1133 TGAAGTGAATATGAGGTATGTGAGCCCTTGAAGTAGAACAGTGAAGAGACCTAGAT 1192
 DB 1090 GGAACGAAAGCAGGGAGTTTCTGCACAGGGGAGTAGGACAGTAAGAGAACCGAGAT 1149


```

Db      35251 ATGCTAGTT 35242
|||||
RESULT 4
AB020590 1823 bp mRNA linear PLN 08-MAY-1999
LOCUS Nicotiana tabacum mRNA for transcription factor NtWRKY2, complete
DEFINITION cds.
ACCESSION AB020590
VERSION 1 GI:4760691
KEYWORDS transcription factor NtWRKY2.
SOURCE Nicotiana tabacum (common tobacco)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
REFERENCE
AUTHORS Shinshi.H., Yamamoto.S. and Suzuki.K.
TITLE Analyses of an elicitor-responsive element and transcription
factors in cultured tobacco cells
JOURNAL Published Only in DataBase (1999)
REFERENCE
AUTHORS Shinshi.H., Yamamoto.S. and Suzuki.K.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-1998) Hideaki Shinshi, National Institute of
Bioscience and Human-Technology, Plant Molecular Biology
Laboratory, Higashi 1-1, Tsukuba, Ibaraki 305-8566, Japan
(E-mail:shinshi@nibh.go.jp, Tel:81-298-54-6071, Fax:81-298-54-6090)
FEATURES
source
1.1823
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Xanthi"
/db_xref="taxon:4097"
/cell_line="XD68"
/cell_type="Suspension culture"
478..1539
/codon_start=1
/product="transcription factor NtWRKY2"
/protein_id="BA077383.1"
/db_xref="GI:4760692"
ENR5YKCTFPNCPTKKKVERSLDQITEIVYKGNHNPKEPOSTRRSSSTASSLRP
TLOYTKLMSQI1SSVGSNATPENSISFGDDHDHDEQSSKSRSGDDNREDEPSK
RWKESSEGLASRGSTVREPRVVTSDIDLDGVWRKYGKQVKGPNPRGY
YKTSFGCPVRKVERASQDIRSVITTEGKGNHDPFAARGSGINRPVAPNITYNNGA
NMAIIRPVSQIPLOSIRPQSPFPTLEMLHKPSNYNGFSGYVNSDSYENQLQDNV
FSRAXDEPRDDMFMTLLC"
1823
polyA_site
/notes="18 A nucleotides"
BASE COUNT 609 a 372 c 333 g 509 t
ORIGIN
Query Match 23.6%; Score 454.2; DB 8; Length 1823;
Best Local Similarity 61.3%; Pred. No. 2.9e-99;
Matches 1070; Conservative 0; Mismatches 473; Indels 202; Gaps 13;
Qy 34 TAGACACTCTGCAAGTGCAAACTCTTTCACCACTTCACCTTCTCCACACACCCCTTCA 93
Db 2 TAGACATCATCTATGAGAAATCAAGTTCTCTTTTCCAACTTCATCTCTTCTTCA 61
Qy 94 TGACCACTCTTCTCTGACTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147
Db 62 TGACTCTCTTCTTCTGACTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 121
Qy 148 CACAGGTGTTGTTCTGAGAGAACTGGCTCTGCTGTTTCCCAAAATTCAGATCCACACCAC 207
Db 122 TTGGTAGATAGTTGACAGAGAACTGGTTCTGGGTTCCCAAGTTTAACTCTCTCCAC 181
Qy 208 CACTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 267
Db 182 CTCCTTCACTACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240

```

```

Qy 268 CTGTTTGTAGCCCTGCTGAGCTTCTTGTACTCGCGGTCTCTCTTAACCTCTTCAACATTC 327
Db 241 CTGGCTTAAGCCCTACTGAACTCTTGACTCTCTCTGTTCTCTTGTCTGCTTCCACATTC 300
Qy 328 TGGCATCTCAACAACTGAGCAATTTGTTGCTCAGAGCTTCAATTGGAAGAGAGTTCAG 387
Db 301 TTCATCTTCAACAAACAGGAGCAT----- 324
Qy 388 GGGGGAATCAGCAAAATTTGTCAAGGAAGAAAGAAAGCTTCTCAAATTTCTCTTTCCAAA 447
Db 325 -----TTCCAGC 331
Qy 448 CCGGATCAGAACCTCTGCTTTCATCCACAGCAACATACCAGTCTTCAAATGTCCACAGTTC 507
Db 332 CCCAAATTGGACAATCTTTTCCATCAGTCTCTCGTCCCAAACTAATCTTGTCCCTTTGA 391
Qy 508 AAACACACAGCAGCATGGAGTTTTCAGAGGCCACGAAACAGGATAATTTTCTCTCAGGAA 567
Db 392 AGAAACAAT---CATGGAATTTCAAGAAATCAAGAAATCAGAATGATCAATCATCTGAAA 448
Qy 568 AGGTATGATGAAAACTGAAAACTCTTCTTCCATGAGAGTTTTCCCTCGAGATTGCTA 627
Db 449 ATGCTAATGGAAGACTGAA-----TTCAACA 475
Qy 628 GTGTCCAAACTTAACCATAGCAATGGGTTCATATCCGATTATGGCAATTAACCCCCACAAT 687
Db 476 CTATGACAGAGCTTGAATAATAATGGTCAGAGCAACCAATATAACAACCAATCATCACA 535
Qy 688 CTCAGACTTTAAGTAGAAGTTCAGATGATGGGTACAAATTGGAGGAAATATGGCCAAAAC 747
Db 536 GTATAAGGGAGCAGAAAGATCAGAAAGATGGGTACAAATTGGAGGAAATATGGCGAAGAA 595
Qy 748 AAGTGAAGGAGTGAAGTCCAAAGAGTTATTACAAATGCACATACCCCAATGGCCCTA 807
Db 596 AAGTGAAGGAGTGAAGTCCAAAGAGTTATTACAAATGCACATACCCCAATGGCCCTA 655
Qy 808 CAAAGAGAGAGTGTAGAGGTCTTTAGATGGCAAAATTTACTGAGATAGTTTATAAGGGA 867
Db 656 CAAAGAGAGAGTGTAGAGGTCTTTAGATGGCAAAATTTACTGAGATAGTTTATAAGGGA 715
Qy 868 CTCATTAACCATCTAAGCTCAAAATATAGAGGAAATCTCATCAAACTCTCTCTCTCTTG 927
Db 716 ATCACAACCATCTCAAAAGCTCTAGTCTACTAGGAGATCTTTCATCAGAGTCTCTCTTAA 775
Qy 928 CAATCCTCATTCAAATTCATCAGAG--ACTGAAATCCAGATCAATCTCTATGCCACACA 985
Db 776 CCGTCCAACTTACAGTACACCAAACTAATGAAATCCAGATCATCAGTT----- 826
Qy 986 TGAAGTGGACAAATGGATTTCAGTCCACCCAGAAAACCTCATCAATATCAATTTGGAGA 1045
Db 827 -----CATATGGTTCAAATGCAACCTGAGAAATCTTCCATTTCAATTTGGTGA 875
Qy 1046 TGATGATTT-----TGAGCAGAGTTTCCAAAAGTCTAAATCAGAGGGGATGAATATGA 1099
Db 876 TGATGATCATGATCATGAGCAAAAGCTCTCAAAGAGTAGGTCAAGAGGAGATCATATGA 935
Qy 1100 TGAAGATGAACCTGATGTCAAAAGATGGAATAATTGAAGGTGAAAATCAGGGTATGTCAGC 1159
Db 936 AGAAGAAGAACCCAGACTCAAAAGATGGAAGAGAAAGCGAAAGTGAAGGTTTATCAGC 995
Qy 1160 CCTGGAAGTAGAACAAGTGAAGAACCTGAGTGTGAGTTGTAGTTTCAGACAAACAGTGCATTGA 1219
Db 996 ACCAGGAGACGACAGTGAAGAACCTGAGTGTGAGTTGTAGTTTCAGACAAACAGTGCATTGA 1055
Qy 1220 TATCTTGTATGATGGCTATAGGTGGAGAAAATACGGGCGAGAAAGTAGTAGTGAAGGGCAATCC 1279
Db 1056 TATTTCTAGATGATGTTATAGATGGAGGAGTATGCTCAAAGAGTAGTGAAGGAAATCC 1115
Qy 1280 AATTCAGAGGTTACTACAGTGCACACACCAGGATGTCCAGTGAAGGAGACAGTGGGA 1339
Db 1116 TAATCCAAAGGGCTACTACAAATGCACAAAGTCCAGGATGTCCAGTAAGAAAACATGTGGA 1175

```

Qy 1340 AAGAGCTCATGACCTAAGGCTGTGATCAACAATTTATGAGGGAAGCACAAACATGA 1399
 Db 1176 ACGAGCATCAAGATATAGATCAGTGATAACACTTATGAGGAATAACACAAACATGA 1235
 Qy 1400 TGTTCCTGACGCCGCTGGCAGTGGCAGCCATTTCTGTGAACAGACCAATGCAACCAATGC 1459
 Db 1236 TGTTCAGCAGCTAGGGGCGAGTGGA-----TAAACCGACCTGTGCGACCAACAT 1286
 Qy 1460 TTCAACACCAACCAACACTGACGACCTTCGTAAGGCTTTGCCAGTGATCACCACAAAG 1519
 Db 1287 TACTTACAACAATGGTCCCAATGCCATGGCCATAAGGCCTT-----CTGTCAATC 1337
 Qy 1520 TGACAAATTCCTTCAGAACCAAGATCACAAAGCACCAAGAGGGCAATCACCTTCAC 1579
 Db 1338 TCAATTTCCATTGCAAGATGATAGGCCACAA-----CAATCTCCCTTAC 1382
 Qy 1580 CCTAGAGATGCTACAAAGTCCAGAACTT---TTGGATTCCTAGGGTTTGGGAATCCAAAT 1636
 Db 1383 ACTAGAGATGTTGCACAGCTAGTAATTAATATGTTTCTCAGGATATGTAATTCAGA 1442
 Qy 1637 GCAATCTAGTGAAACAGACGAACTATCTGACAATGTTTCTCTCCAGGACCAAGGA 1696
 Db 1443 GGATTCATATCAAAAC-----CAACTTCAGGACAATTAATGTAATTTCCAGAGCAAGGA 1496
 Qy 1697 GGACCTAGAGATGACATGTTCTTGTAGTCTCTACTATGCTGAAGGAATTTTTCCTCC 1756
 Db 1497 CGACCTCGAGATGACATGTTATGAGACATGCTCTGCTGATGTAATTCGATTCCTGGTC 1556
 Qy 1757 CTTT 1761
 Db 1557 ATCT 1561

RESULT 5
 AB063573
 LOCUS Nicotiana tabacum NtWRKY-6 mRNA for WRKY DNA-binding protein, complete cds.
 AB063573.1 GI:14530680
 ORGANISM Nicotiana tabacum (common tobacco)
 Nicotiana tabacum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.
 1 (sites)
 Maeo.K., Hayashi,S., Kojima-Suzuki,H., Morikami,A. and Nakamura,K.
 Role of conserved residues of the WRKY domain in the DNA-binding activity of tobacco WRKY family proteins
 Unpublished
 2 (bases 1 to 1988)
 Maeo,K., Hayashi,S., Kojima-Suzuki,H., Morikami,A. and Nakamura,K.
 Direct Submission
 Submitted (21-JUN-2001) Kenichiro Maeo, Nagoya university, Graduate school of bioagricultural sciences; Chikusa-ku, Nagoya, Aichi 464-8601, Japan (E-mail:maeo@agr.nagoya-u.ac.jp, Tel:81-052-789-4096, Fax:81-052-789-4095)
 Location/Qualifiers
 1..1988
 /organism="Nicotiana tabacum"
 /mol_type="mRNA"
 /cultivar="Petit Havana SRI"
 /db_xref="taxon:4097"
 1..1988
 /gene="NtWRKY-6"
 48..1727
 /gene="NtWRKY-6"
 /note="tobacco WRKY family proteins"
 /codon_start=1
 /product="WRKY DNA-binding protein"
 /protein_id="BAB61053.1"
 /db_xref="GI:14530681"

/translation="MASGGNNVTFMNSFTSNYSFSSFDLLSDNDNNNNNNRRNNSS
 MNQENSLNMFSDQRMHQNDKVPKFKSPFSLPMISSSPSAATSSSPSYLAFPPSL
 PSVLDSPLVFNNTLPSPITTSFGLNLSKEDDSRIIDFSPQSRQSNPPPVHTVTPSQY
 RNSLDLMTROQHANOQEFSTAKTTGVKEVVPISQFQSRQSNPPPVHTVTPSQY
 VNRQAEQGNWRKYGQKQVQSENPRSYKCTFPNCTPKKVERNLGDGHIETIVYKG
 NNHPKPOSTRRSSSSQNLAYSLNIDITNQAFELDNAQRDSFAGTDSNSASFGDED
 IDQGFVSKSGEDDGNPEAKGKGNENEVISSASRTVREPRIVVQITSDIDLDDG
 YWRKYGQKVPKGNPRSYKCTFTGCPVKRHSRDLRAVITTYEGKNHDPVA
 ARGSGSYAMNKPPSGNSNMPVPRPMLANNQNMNFDTFFNTRVQITQNOQPP1
 TLQMLQSSSSSYSGFDTSYSGYMNQMSNIKPTKEPKDDFFSSPLN"
 BASE COUNT 671 a 388 c 366 g 563 t
 ORIGIN
 Query Match 22.3%; Score 430.6; DB 8; Length 1988;
 Best Local Similarity 61.0%; Pred. No. 1.5e-93;
 Matches 948; Conservative 0; Mismatches 499; Indels 108; Gaps 11;
 Qy 179 TGTGTTCCTCAATCAAGTCCACACCACTTCTCTGCG-----TCTCTCTCCCC 232
 Db 248 TGAAGTTCCAAAGTTTAAAGTCATTTCCACCTTGTCTTTGCCAATGATTTCATCTCATC 307
 Qy 233 TCCATTTCTCTCTTCTTACTTTTGCTATTCTCTCTGTTTGAGCCCTGCTGAGCTTCT 292
 Db 308 ACCAGCTTCTCTCTTCTTATCTTCTCTCTCTCTTCTTAAAGTCCATCTGACTTTT 367
 Qy 293 TGACTCGCGGTTCTCTTAACTCTTCCAACTTCTGCCATCTCTCAACCACTCGAGCATTT 352
 Db 368 GGAATCACCAGTTTGTGTTTAACTTCCAACTTCTCTTCCATCTCTCAACTACAGGAGTTT 427
 Qy 353 TGTGTCTCAGAGCTTCAATTTGGAAGAGAGCTTTCAGGGGGGATCAGCAATTTGTCAAGGA 412
 Db 428 TGGTAATTTGAATTC-----AAGG 448
 Qy 413 AGAAGACAAAGCTTCTCAAAATTTCTTTCGAAACCCGATCAGGACCTCTCTCTCATC 472
 Db 449 GGATGACTCAAGGATTTCTGATTTCTTTCGAAAGTAGGGCTG---CTACTTTCATCATC 505
 Qy 473 CACAGCAACATACCACTCTTCAAAATGTCAGTTTCAACACACACAGCCCTGAGTTTTC 532
 Db 506 AATGTTTCACTTCTTCTCTCAAGAACTCATTTGGAAGACTTAATGACA---AGGCAACA 562
 Qy 533 GGAGGCCACGAAACAGGATAATTTTCTCAGAAAGGGTATGATGAAACTGAAACTC 592
 Db 563 ACATGCCAACACGAAATGAAATTTCTCCACTGCAAAACTTACAGGGGTGAAATCAGAACT 622
 Qy 593 TTTCTTCATGCGAGGTTTTCCTCCTGAGATTGCTAGTGTCCAACTAACCATAGCAATGG 652
 Db 623 GGTTCCAATTCAAAGCTTCTCCAGGAG-----AAGATGCA 658
 Qy 653 GTTTCATCCGATTTATGGCAATTTACCCCCCACTCTCAGACTTTTAAGTAGA---AGGTC 709
 Db 659 GAGTAATCCACTCCAGTACATTTACACTCAACCATCTCAATATGTTAGAGAACAGAAAGC 718
 Qy 710 AGATGATGGGTCAATTTGAGGAAATATGCGCAAAACCAAGTGAAGGAAAGTGAATAATCC 769
 Db 719 AGAAGATGGATACAAATTTGAGGAAATATGGCAGAGCAAGTGAAGGAAAGCAGAGATCC 778
 Qy 770 AAGAAGTTTATCAAAATGCATATCCCAATTCCTCTTACAAAGAGAGGTTTCAGAGGTC 829
 Db 779 GAGAAGCTATTACAAGTGACATTTCCCTACTCTCTTACAAAGAGAGGTTTGAAGGAA 838
 Qy 830 TTTAGATGGACAAATTTACTGAGATAGTTTATAGGGGTACTCTATAACCATCTTAAGCCTCA 889
 Db 839 TTTGATGGACACATTTACTGAGATAGTTTATAGGGGAACCATATCATCTCAAGGCTCA 898
 Qy 890 AATATCTAGGAGAACTCATCAAACTCTCTTCTTTGCAATCCCTCATTTCAATTCAT 949
 Db 899 GTCCACTAGAAGA-----TCGTTCTTCAAGTTCGATTCATAAACCTTCTTCTTACTC 946
 Qy 950 CAGAACTGAATCCAGATCAATCTTATGCCACATGGAAGGAGCAATGATTCAGC 1009
 Db 947 CAACCTTGGATATAACAAATCAGCAACCGCGTTTCTTGATATGCTCAAGAGATTCTT 1006


```

QY 1179 AGAGAACCTAGAGTTCTAGTTCAGACAAACAGTGACATTCATCTTCATGATGCTAT 1238
Db 1203 CGCGACCAAGATTCTAGTCCAAACAACTAGTGATATAGATATCTTCATGATGCTAT 1262
QY 1239 AGGTGAGAAATATACCGGAGAAAGTAGTGAAGGCAATCCAAATCAAGAGTACTATAC 1298
Db 1263 AGGTGAGAAATATAGCAGAGAGTGTGTTAAAGGCAATCCAAATCCTAGGAGCTACTAC 1322
QY 1299 AAGTGACACACCCAGGATGTCAGTGAGGAGCAGCTGGAAGAGCCTCACATGACCTA 1358
Db 1323 AATGTACTCAAGTTGGCTGTCCGGTGAGAAACACAGTGAGAGCATCTCATGATTTG 1382
QY 1359 AGGGCTGTGATCAACACTTATGAGGAAAGCAACCAATCATGATCTTCTGAGCCCGTGGC 1418
Db 1383 AGGGCAGTGATCAACACTTACGAAGGAAACATAACCATGATGTCTCTGCACACCGCGC 1442
QY 1419 AGTGGAGCATTCTGTGAACAGACCAATGCCAAACAAATGCTTCAAAACACACCAACT 1478
Db 1443 AGTGGCAGTTATGCAGTAAATAGACCATCTGTATAACACCGC-----CACCACTAGT 1493
QY 1479 GCAGGCACCTTCCGTAGGCTCTTCCAGTGATCCACCAAGTGACAAATCCCTTCAGAAC 1538
Db 1494 GTCTCCACCGCTATAAGC-----CAACCAAAATTAATTAACCTCTTTCAGAAC 1544
QY 1539 CAAAGATACCAAGCACCACAGAAAGGCAATCACCTTACCCCTAGAGATGCTACAAAGT 1598
Db 1545 ACTAGAGCACA---ACTGCAATGCGCAAGCACCAATTTACACTAGAAATGCTGACGGT 1601
QY 1599 CCAGGAAGTTTTCGATCTCAGGTTTGGGAATCCAAATGCAATCTTACGTGAACACGACG 1658
Db 1602 CCAGGAAGTACGAGTTCTCAGGTTTGGAACTCACTAAATTAATTAACAGATCAACCAA 1661
QY 1659 CAATCTGACAAATGTTTCTCTCCAGACCAAGGAGGAGGCTAGAGATGACATGTTTC 1718
Db 1662 AATCAACAGCAAGTGACAGATTTCTCCACAGCCAGGAGGAGGCTGATGTTGACTCATTT 1721
QY 1719 CTTGAGTCTTACTATGCTGA 1739
Db 1722 TTTGACTCTTTTGGCCTGA 1742

RESULT 8
PCU48831 1800 bp mRNA linear PLN 19-NOV-1996
LOCUS Petroselinum crispum DNA-binding protein WRKY1 mRNA, complete cds.
DEFINITION U48831
ACCESSION U48831.1 GI:1431871
VERSION 1
KEYWORDS Petroselinum crispum (parsley)
SOURCE Petroselinum crispum
ORGANISM Petroselinum crispum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; campanulids; Apiales; Apiaceae; Petroselinum.
REFERENCE 1 (bases 1 to 1800)
AUTHORS Rushon,P.J., Torres,J.T., Parniske,M., Wernert,P., Hahlbrock,K.
and Somasich,I.E.
TITLE Interaction of elicitor-induced DNA-binding proteins with elicitor
response elements in the promoters of parsley PRI genes
JOURNAL EMBO J. 15 (20), 5690-5700 (1996)
MEDLINE 97051827
PUBMED 8896462
REFERENCE 2 (bases 1 to 1800)
AUTHORS Rushon,P.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-1996) Paul Rushton, Dept of Biochemistry,
Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne Weg 10,
Koeln 50829, Germany
FEATURES
source Location/Qualifiers
1..1800
/organism="Petroselinum crispum"
/mol_type="mRNA"
/db_xref="taxon:4043"

```

52..1596
 /note="contains two WRKY domains; WRKY-type DNA-binding protein; sequence-specific DNA-binding protein"
 /codon_start=1
 /product="WRKY1"
 /protein_id="AAC49527.1"
 /db_xref="GI:1431872"
 /translation="MSSSLGLDILAQPNNDFFGSMNGFNQKTSFANQSLPFPSPPPVPS
 PSSYFSLDPSLQNNNTIVSSGNGKTNNAQSFKEENQNFSDPFAOTRPMNQAAN
 LI PANTNIVEDSLKRGKGNFEBPAKNDPLMDNASVTSDIATLQRIPEMTMNAQAN
 MQSNAALQSNLNNYAQSQSQTNRDQSKLDDGYNRKYGOKQVKGSENPRSYVYKCTY
 LMCPTKPKVETFDGHTEIVYKGNHNPKEPOSTKRSSQSYQNSIGTNPESLLENG
 RSEPTVTTPENSLSFGEDDLFEQGNMKNPDDDDNEPDSKRWKGYTESNEPWSLLSR
 TVREPRIVVITTDIILDGVWRKYGQVKGPNPRSYKCTQVGCPRKRVHVERA
 SHDLRAVITTYEGKHNDVPAKRGSGSYPAVNRPSDNTTSAPTAIRPTNYLNPQNP
 RAOPANGQAPFTLEMLQRPSPRYEFGFTNTSNTYAINQNAQSGQSFSTAKDPEVDSE
 FDSFLA"

misc_feature 340..357
 /note="encodes serine/threonine-rich region"
 misc_feature 493..612
 /note="encodes glutamine-rich region"
 misc_feature 627..789
 /note="encodes WRKY domain and a putative DNA-binding domain and zinc finger domain"
 misc_feature 970..993
 /note="encodes acidic domain"
 misc_feature 997..1008
 /note="encodes basic domain"
 misc_feature 1093..1113
 /note="encodes acidic domain"
 misc_feature 1111..1278
 /note="encodes WRKY domain"

BASE COUNT 610 a 379 c 334 g 477 t
 ORIGIN

Query Match 19.7%; Score 379; DB 8; Length 1800;
 Best Local Similarity 65.9%; Pred. No. 5.1e-81;
 Matches 686; Conservative 0; Mismatches 295; Indels 60; Gaps 7;

QY 702 AGAAGTCAGATGATGGGTACAAATTCGAGGAATATGGCCAAAACAAAGTGAAGGAAGT 761
 Db 613 AGTAAATTCGACGATGGATACAAATTCGAGGAATATGGGACAGAACGAGTTAAGGAAGT 672
 QY 762 GAAATTCGAGGAATGATTAATTCACAAATGCACATACCCCAATTCGCCCTACAAAGAAAGAGTT 821
 Db 673 GAGATTCGAGGAATGATTAATTCACAAATGCACATATCTTAATTCCTTACTAAGAAAAGGTG 732
 QY 822 GAGAGTCTTTAGATGGACAAATTAATTCGAGATAGTTTATAAGGGTACTATACCAATCCCT 881
 Db 733 GAAACAACTTTTATGGACACATAACTGAAATTTGTATACAAAGGCAATCATACCATCCA 792
 QY 882 AAGCCTCAAAATTAATTCAGGAGAACTCATCAAACTCTCTCTCTCTTCTGCAATCCCTCATTC 941
 Db 793 AAGCCTCAATCAACGAAACGA-----TCACTTCACAGTCTTACCAAGAATTC 840
 QY 942 AATTCATCAGAACTGAAATCCAGATCAATCTATGCACACATGCAAGAGTGACAAATG 1001
 Db 841 ATAGGTACCATGCCAGAA-----GTTCTTACTAGAAATGCGCGTCA 885
 QY 1002 GATTCAGTGCACCCGAGAAATCTCATCAATATCAATTTGGAGATGATGATTTTGAGCAG 1061
 Db 886 GAACCAAGTGCACACCCCTGAAATTTCTTCGCTTTCTTTTGGCGAGGATGACCTCTTTGAA 945
 QY 1062 AGTTCCCAAAAGTGAATTAATTCAGGAGGGATGATATGATGAGATCAACTGATGCCAAA 1121
 Db 946 CAAGGATTCATGAATAAACACGAGATG-----ATGATGGAATGAACCTGATTTAAA 999
 QY 1122 AGATGAAAAATGAAGTGA---AAATGAGGGTATCTCAGCCCTCAGAGTGAACAGATG 1178
 Db 1000 AGATGGAAGGTGAATATGAGAGTAAATGAACCAATGTCATCTCTGGGAGTCAATCTGTC 1059
 QY 1179 AGAGAACCTAGAGTTCTAGTTCAGACAAACAGTGAATTCATCTTGTATGATGCTAT 1238

```

Db      1060 CGCGAACCAAGATTGTAGTCCAAACAACTAGTGTATAGATATCTTTGATGATGTTAT 1119
Qy      1239 AGTGGAGAAAAATACGGGCGAAGAGTAGTGAAGGCGAATCCAAATCCAGGAGTTACTAC 1298
Db      1120 AGTGGAGAAAGTAGTGCACAGAAAGTTGTGTTAAAGGCAATCCAAATCCTAGGAGTACTAC 1179
Qy      1299 AAGTGCACACACCCAGGATGTCAGTGCAGGAGCAAGCAGTGAAGAGAGCTCACAATGACCTA 1358
Db      1180 AATGTGTACAAAGTTGGCTGTCCAGTGCAGGAGCAAGCAGTGAAGAGAGCTCACAATGACCTA 1239
Qy      1359 AGGCTGTGATCAAACTTATGAGGGAAGCAACCAACCATGATGTTCTCGAGCCCGTGGC 1418
Db      1240 AGGCACTGATCACCACTTACGAGGGAACACATAAACCAAGTGTCTCTGCCCCGAGGC 1299
Qy      1419 AGTGGCAGCCATCTCTGTAACAGACCAATGCGCAAAACAAATGCTTCAAAACCAACCAACACT 1478
Db      1300 AGTGGCAGTTATCTCTG-----CAGTGAACAGGCGCATCTGATACACACCACCACT 1347
Qy      1479 GCAGCCACTTCGTAAGGCTTTGCCAGTGTCCACCAAGTGACATTCCTTCCAGAAC 1538
Db      1348 GCTCTACGGCTATAAGGCCAACACCA-----AATTACTTAAACCTCTTTCAGAAC 1398
Qy      1539 CAAAGATCAACAAGCACCACAGAGGCAATCACCTTCCACCTTAGAGATGCTACAAAGT 1598
Db      1399 CCAAGAGCACA--ACCTGCCAATGGTCAAGCACCAATTACACTAGAAATGCTGCAACGT 1455
Qy      1599 CCAGGAAGTTTGGATTTCTCAGGTTTGGGAATCAATGCAATCTTACGTGAACAGCAG 1558
Db      1456 CCAGGAAGTACGAGTTCTCAGGTTTACAAACACATCCCAATCTTACGGCGATCAACCA 1515
Qy      1659 CAACTATCTGCATGTTTCTCTCCAGGACCAAGGAGGCGCTAGAGATGACATGCTTC 1718
Db      1516 AATCAACAGGCGCAGTGACAGTCTCTCCAGGCAAGGACGAGCGCTGATGTTGACTCGTTT 1575
Qy      1719 CTTGAGTCTCTACTATGCTGA 1739
Db      1576 TTTGACTCTTTTATGCTGA 1596

RESULT 9
AB063575
LOCUS   Nicotiana tabacum NtWRKY-8 mRNA for WRKY DNA-binding protein,
DEFINITION partial cds.
ACCESSION AB063575
VERSION   AB063575.1 GI:14530684
KEYWORDS
SOURCE    Nicotiana tabacum (common tobacco)
ORGANISM  Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
1 (sites)
Mao,K., Hayashi,S., Kojima-Suzuki,H., Morikami,A. and Nakamura,K.
Role of conserved residues of the WRKY domain in the DNA-binding
activity of tobacco WRKY family proteins
Unpublished
2 (bases 1 to 1417)
Mao,K., Hayashi,S., Kojima-Suzuki,H., Morikami,A. and Nakamura,K.
Direct Submission
Submitted (21-JUN-2001) Kenichiro Mao, Nagoya university, Graduate
school of biocultural sciences; Chikusa-ku, Nagoya, Aichi
464-8601, Japan (E-mail:mao@agr.nagoya-u.ac.jp,
Tel:81-052-789-4096, Fax:81-052-789-4095)
1. 1417
Location/Qualifiers
1. 1417
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Petit Havana SRI"
/db_xref="taxon:4097"
1. 1417
/gene="NtWRKY-8"
&lt2. .1138

gene
CDS

```

```

/gene="NtWRKY-8"
/note="tobacco WRKY family proteins"
/codon_start=1
/product="WRKY DNA-binding protein"
/protein_id="BAB61055.1"
/db_xref="GI:14530684"
/translation="KNYSDFSPQPOFASVPSQTNPLGKSMYQESRKNDENAN
GTSLEQLKNGQSNQYKSSRSDGYNWKYQKQVGSNPSRYKCTPPNCTPK
KVEELDDQITETIVYKGNHHPKPTQSRSSSLAIQYNTQTNIEIPHQSTPNS
ISFGDLDHEKSRSGDDPEEEDPSKEDPKRWKRESEGLSGASTVREPRVVV
QTSDIDLDGDRWKYQKQVKNPNPSRYKCTSPCPYRKVHVERASQDIRSVIT
SYGNKHNDVPAARSAINRPVATPIRPSVTSQIPLPQSPFLEMLHKP
SYNGHPSGYATSEDSYENQLDNNVPSRAKDEPRDDMFMTLLC"
BASE COUNT 513 a 269 c 275 g 360 t
ORIGIN
Query Match 18.3%; Score 353.4; DB 8; Length 1417;
Best Local Similarity 68.7%; Pred. No. 8.5e-75;
Matches 571; Conservative 0; Mismatches 206; Indels 54; Gaps 4;
Qy      684 CAATCTCAGACTTTAAGTAGAAGGTCAGATGAGGTACAAATTTGGAGGAAATATGCCAA 743
Db      179 CAATATAAAGCAATCATCAAGATCAGAAGATGGGTACAAATGGAGGAAATATGGCAG 238
Qy      744 AAACAAAGTGAAGGAAAGTGAATAATCCAAAGAGTTATTACAAATGCACATACCCCAATTGC 803
Db      239 AAACAAAGTGAAGGTTAGTGAATAATCCAAAGAGTTACTATAAATGTACTTTCCCAATTGT 298
Qy      804 CTTACAAAGAAAGGTTGAGAGGCTTTAGATGGAATAATTTACTGAGATAGTTTATAAG 863
Db      299 CCCACAAAGAAAGTGAAGAGGTTTGTAGATGGGCAAAATTAATCGAAATTTGTACAAG 358
Qy      864 GGTACTCATTAACCATCTTAAGCCTCAAAATCTAGGAGAACTCATCAAACTCTCTTCT 923
Db      359 GGTAAATCAACCATCCAAAGCGGACTCAGTCTACTAG-----GAGATCTCTCT 409
Qy      924 CTTGCAATCCCTCATTTCAAATTTCCATCAGAACTGAAATCCCAGATCAATCAATATCAATTGGA 1043
Db      410 TTAGCCATCAACCTTTACAAATACAAATCTAATGAAATCCAGATCATCAGT----- 461
Qy      984 CATGGAAGTGGACAAATGGATTCAGTGCACCCAGAGAAACTCATCAATATCAATGGA 1043
Db      462 -----CAACACCTGAGAAATTCGTCATTTTCATTTGGG 493
Qy      1044 GATGATGATTTTCAGCAGAGTTCCCAAA-----AGTGAATCAGAGGGGATGAATAT 1097
Db      494 GATGATGATCATGAGAGAGTAGTCAAGAGGAGATGATTTCCATGAGAGAGAACCA 553
Qy      1098 GATGAAGATGAACCTTGATGCCAAAAGATGAAAATTTGAAGGTGAAAATGAGGGTATGTCA 1157
Db      554 GACTCAAAAGAACCAAGACCCCAAAAGATGAAAAGAGAAAGCGAAAGTGAAGTTTATCA 613
Qy      1158 GCCCCTGGAAAGTAGAAGTGAAGAACCTTAGAGTTGTAGTTTCAGACCAACAGTGACATT 1217
Db      614 GCACCAAGAGCAGAACAGTGAAGAACCTTAGAGTTGTAGTTTCAAACTCAAGTGATATT 673
Qy      1218 GATATCTTTCATGATGCTTAGTGTGAGAAAATACGGGAGAAAATGAGTGAAGGGCAAT 1277
Db      674 GATATTTAGATGATGATGATGATGAGGAGATGATGTTCAAAAGTAGTCAAAAGGAAT 733
Qy      1278 CCAAAATCCAAAGGAGTTACTACAAAGTGCACACCCAGGATGTCCAGTGAGGAAGCAGTG 1337
Db      734 CTTAATCCAAAGGAGTACTACAAATGCACAAAGTCCAGATGTCCAGTGAAGAAACATGTG 793
Qy      1338 GAAAGAGCCTCAGATGACCTTAAGGGCTGTGATCACAATTTATGAGGAAAGCACAACCAT 1397
Db      794 GAACGAGCATCAAAAGACATGAATCAGTGTATGAACCACTTATGAAGGAAACCAACAC 853
Qy      1398 GATGTTCTCCAGCCCTGGCAGTG---GCAGGCAATTTCTGTGAACACACCAATGCCAAC 1454
Db      854 GATGTTCCAGCAGCTAGGGCAGTGCCATCAACCGACCTGTGCACCAACCATTTACTTAC 913
Qy      1455 AATGCTTTCAAAACCAACCAACCACTGCAGCCACTTCGTAAGGCTCTTTGCCA 1505

```



```
Db 914 AACAAATCCATCCCAAGGCTTCTGTACATCTCAAAATTCATTCGCA 964
|||||
RESULT 10
AX412575 1539 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 339 from Patent WO0222675.
ACCESSION AX412575
VERSION AX412575.1 GI:21445033
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Glazebrook, J., Wang, X., Dangel, J. L., Eulgem, T. and Zhu, T.
TITLE Plant genes, the expression of which are altered by pathogen
infection
JOURNAL Syngenta Participations AG (CH); UNIVERSITY OF NORTH CAROLINA AT
CHAPEL HILL (US); Glazebrook, Jan (US); Wang, Xun (US); Dangel,
Jeffrey L. (US); Eulgem, Thomas (US)
FEATURES
source
Location/Qualifiers
1..1539
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 506 a 355 c 309 g 369 t
ORIGIN
Query Match 16.0%; Score 308.2; DB 6; Length 1539;
Best Local Similarity 67.1%; Pred. No. 7.7e-64;
Matches 493; Conservative 0; Mismatches 218; Indels 24; Gaps 3;
QY 710 AGATGATGGGTACAAATGGAGGAATATGCGCAAAACAAAGTGAAGGAAGTGAAATCC 769
DB 528 AGAGGATGGTTACAAATGGAGGAATATGCGCAAAACAAAGTGAAGGAAGTGAAATCC 587
QY 770 AAGAGTTATTAACAATGCATACCCCAATTCGCCCTTACAAAGAGAGTTGAGAGTCC 829
DB 588 TCGGAGTTACTATAAGTGTACTTTCCCTTAATTTGCCAAGGAAGGAAGTGAGAGATC 647
QY 830 TTTAGATGGCAAAATTAAGTGTACTTTATAGGCTCTCTAAGCCCTTCAAGCTCA 889
DB 648 TTTGGAAGGTGAGATCAGAGATTTGTGTATAAGGAAGCAACCATCTTAAACCTCA 707
QY 890 AAATCTAGGAGAACTCATCAAACTCTCTCTTCTTCAATCCCTCATTCAAATTCAT 949
DB 708 GTCTACTAGAAGATCTTCTCGTCTCTTCGACITTTTCATTACGCTGTGTACATGCCAG 767
QY 950 CAGAACTGAATCC-----CAGATCAATCTCTATGCCACATCGGAAGTGACA 997
DB 768 TTTGGATCATAATCGTCAAGCTTCTCTGATCAGCTTAATTCCTAATAATAGCTTTTCATCA 827
QY 998 AATGGATTACGTCGCCACCCAGAA-----AACTCATCAATCAATTCGAGATGA 1048
DB 828 GTCTGATTCCTTTGGGATGCAACAAGAGGATAATACTACTTCTGTGTTGAGTACGA 887
QY 1049 TGATTTTGAGCAGAGTTTCCCAAAAGTGAATCAGAGGGGATGAATATGATGAAGATGA 1108
DB 888 TGAGTTCGAACAGGCTC---ATCGATTTGTCAGCAGAGAGCAAGATTTGTGGAGTGA 944
QY 1109 ACCTGATGCCAAAGATGGAATAATGAGGTGGAATAATGAGGGTATGTCAGCCCTCGAAG 1168
DB 945 ACCTGAAGCAAGAGATGGAAGGGGACAATGAAACAAATGGTGGGAATGGTGGTGAAG 1004
QY 1169 TAGACAGTGAAGAACTAGAGTTGATTTGATTCAGACAAACAGTGCATTTGATTCCTTGA 1228
DB 1005 CAAGACAGTGAAGACCGGGAATCGTAGTGCAGACAAACAGGATGATATTCATCTTTGA 1064
QY 1229 TGATGGCTATAGGTGGGAAATAACGGGCAGAAAGTAGTAGTGAAGGGCAATCCAAATCCAAAG 1288
|||||
Db 1065 CGACGGTTACAGATGGAGAAATACGCGCCAGAAAGTCGTTAAGGGAACCCAAATCCAAG 1124
|||||
QY 1289 GAGTTACTCAAGTGCACACACCAGCATGTCAGTGAGGAACACGTCGTGGAAGAGCCTC 1348
|||||
DB 1125 AAGCTACTCAAGTGCACACACCATCGTGTGTCAGTGAGGAACACATGTGGAGAGATC 1184
|||||
QY 1349 ACATGACCTAAGGGCTGTGATCACAATTTATAGGGGAAAGCACAACCATGATTTCTGTC 1408
|||||
DB 1185 ACACGACATGAGAGCAGTAATCACAACCTACGAAAGGGAACACACACCATGTTCTCTGC 1244
|||||
QY 1409 AGCCCTGGCAGTGG 1423
|||||
DB 1245 AGCTCGTGGTAGCGG 1259
|||||
RESULT 11
AX412905 1539 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 669 from Patent WO0222675.
ACCESSION AX412905
VERSION AX412905.1 GI:21445363
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Glazebrook, J., Wang, X., Dangel, J. L., Eulgem, T. and Zhu, T.
TITLE Plant genes, the expression of which are altered by pathogen
infection
JOURNAL Syngenta Participations AG (CH); UNIVERSITY OF NORTH CAROLINA AT
CHAPEL HILL (US); Glazebrook, Jan (US); Wang, Xun (US); Dangel,
Jeffrey L. (US); Eulgem, Thomas (US)
FEATURES
source
Location/Qualifiers
1..1539
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 506 a 355 c 309 g 369 t
ORIGIN
Query Match 16.0%; Score 308.2; DB 6; Length 1539;
Best Local Similarity 67.1%; Pred. No. 7.7e-64;
Matches 493; Conservative 0; Mismatches 218; Indels 24; Gaps 3;
QY 710 AGATGATGGGTACAAATGGAGGAATATGCGCAAAACAAAGTGAAGGAAGTGAAATCC 769
DB 528 AGAGGATGGTTACAAATGGAGGAATATGCGCAAAACAAAGTGAAGGAAGTGAGATCC 587
QY 770 AAGAGTTATTAACAATGCATACCCCAATTCGCCCTTACAAAGAGAGTTGAGAGTCC 829
DB 588 TCGGAGTTACTATAAGTGTACTTTCCCTTAATTTGCCAAGGAAGGAAGTGAGAGATC 647
QY 830 TTTAGATGGCAAAATTAAGTGTACTTTATAGGCTCTCTAAGCCCTTCAAGCTCA 889
DB 648 TTTGGAAGGTGAGATCAGAGATTTGTGTATAAGGAAGCAACCATCTTAAACCTCA 707
QY 890 AAATCTAGGAGAACTCATCAAACTCTCTCTTCTTCAATCCCTCATTCAAATTCAT 949
DB 708 GTCTACTAGAAGATCTTCTCGTCTCTTCGACITTTTCATTACGCTGTGTACATGCCAG 767
QY 950 CAGAACTGAATCC-----CAGATCAATCTCTATGCCACATCGGAAGTGACA 997
DB 768 TTTGGATCATAATCGTCAAGCTTCTCTGATCAGCTTAATTCCTAATAATAGCTTTTCATCA 827
QY 998 AATGGATTACGTCGCCACCCAGAA-----AACTCATCAATCAATTCGAGATGA 1048
DB 828 GTCTGATTCCTTTGGGATGCAACAAGAGGATAATACTACTTCTGTGTTGAGTACGA 887
QY 1049 TGATTTTGAGCAGAGTTTCCCAAAAGTGAATCAGAGGGGATGAATATGATGAAGATGA 1108
DB 888 TGAGTTCGAACAGGCTC---ATCGATTTGTCAGCAGAGAGCAAGATTTGTGGAGTGA 944
QY 1109 ACCTGATGCCAAAGATGGAATAATGAGGTGGAATAATGAGGGTATGTCAGCCCTCGAAG 1168
DB 945 ACCTGAAGCAAGAGATGGAAGGGGACAATGAAACAAATGGTGGGAATGGTGGTGAAG 1004
QY 1169 TAGACAGTGAAGAACTAGAGTTGATTTGATTCAGACAAACAGTGCATTTGATTCCTTGA 1228
DB 1005 CAAGACAGTGAAGACCGGGAATCGTAGTGCAGACAAACAGGATGATATTCATCTTTGA 1064
QY 1229 TGATGGCTATAGGTGGGAAATAACGGGCAGAAAGTAGTAGTGAAGGGCAATCCAAATCCAAAG 1288
|||||
```



```

Db      888  TGAGTTGGAACAGGCTC---ATCGATTGTCAGCAGACGAGAGAGATTGTGGAGTGA 944
Qy      1109  ACCTGATGCCAAAGATGGAATAATTGAAGGTGAAATAGGGTATGTCAGCCCTGGAAG 1168
Db      945  ACCTGAGCAAGAGATGGAAGGGGCAATGAAACAAATGGTGGGAATGGTGGGAAG 1004
Qy      1169  TAGAACAGTAGAGAACTTAGAGTTGTAGTTCAGACAAACAGTGACATTTGATATCTCTGA 1228
Db      1005  CAAGACAGTAGAGAGCGGAGAAATCGTAGTCAGACAAACGAGTGATATTGACATTTCTTGA 1064
Qy      1229  TGATGGCTATAGGTGGGAGAAATACGGCGCAGAAAGTAGTGAAGGGCAATCCAAATCCAAG 1288
Db      1065  CGACGGTTACAGTAGAGAAATACGGCGCAGAAAGTAGTGAAGGGCAATCCAAATCCAAG 1124
Qy      1289  GAGTTACTACAGTGACACACCCAGGATGTCAGTGAAGGACACCAACCATGATGTTCTCTGC 1348
Db      1125  AAGCTACTACAGTGACACACCCAGGATGTCAGTGAAGGACACCAACCATGATGTTCTCTGC 1184
Qy      1349  ACATGACCTAAGGCTGTGATCAACCTTATGAGGGAAGACCAACCATGATGTTCTCTGC 1408
Db      1185  ACAGACATGAGAGCAGTAATCACACCTAGAAAGGGAACACCAACCATGATGTTCTCTGC 1244
Qy      1409  AGCCCGTGGCAGTGG 1423
Db      1245  AGCTCGTGGTAGCGG 1259

```

```

RESULT 12
AX506109  AX506109 1539 bp DNA linear PAT 27-SEP-2002
LOCUS
DEFINITION Sequence 804 from Patent WO0216655.
ACCESSION AX506109
VERSION AX506109.1 GI:23387346
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Harper, J. P., Krups, J., Wang, X. and Zhu, T.
TITLE Stress-regulated genes of plants, transgenic plants containing
same, and methods of use
JOURNAL Patent: WO 0216655-A 804 28-FEB-2002;
The Scripps Research Institute (US) ; Syngenta Participations AG
(CH)
FEATURES
source
Location/Qualifiers
1..1539
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 506 a 355 c 309 g 369 t
ORIGIN
Query Match 16.0%; Score 308.2; DB 6; Length 1539;
Best Local Similarity 67.1%; Pred. No. 7.7e-64;
Matches 493; Conservative 0; Mismatches 218; Indels 24; Gaps 3;

```

```

Qy      710  AGATGATGGGTACAAATGGAGGAATATGGCCAAACAAAGTGAAGGAGTGAATCC 769
Db      528  AGAGATGGTTACAAATGGAGGAATATGGCCAAACAAAGTGAAGGAGTGAATCC 587
Qy      770  AAGAAGTTATTACAAATGCATACCCCAATTCGCCCTACAAGAGAGAGTTGAGAGTCC 829
Db      588  TCGAGTTACTATAAGTGTACTTCCCTAATTTGTCACGAGAGAGAGAGTGGAGAGATC 647
Qy      830  TTTAGATGGCAAAATTACTAGATAGTTTATAAGGGTACTCATTAACCATCTTAAGCCTCA 889
Db      648  TTTGGAAGGTGAGATCAAGAGATTGTGTATAAAGGAGGCCAACCATCTCTAAACCTCA 707
Qy      890  AAATACTAGGAGAACTCATCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT 949

```

```

Db      708  GTCTACTAGAGATCTTCTTCTGCTTTCTTCGACTTTTCTTCAATTCAGCTGTGTACAATGCCAG 767
Qy      950  CAGAACTGAAATCC-----CAGATCAATCTATGCCACACATGGAAGTGGACA 997
Db      768  TTTGGATCAATAATCGTCAAGCTTCTTCTGATCAGCCTAATTCATAATAGTCTTCATCA 827
Qy      998  AATGGATTTCAGCTGCCACCCAGAA-----AATCATCAATATCAATTTGGAGATGA 1048
Db      828  GTCCTGATTCCTTTGGGATGCAACAGAGGATAATACTACTTCTGATTTCTGTTGGTGACGA 887
Qy      1049  TGATTTTGAGCAGAGTTCCAAAAGTGTAAATCAGAGGGGATGAATATGATGAAGATGA 1108
Db      888  TGAGTTTCGAACAGGCTC---ATCGATTGTCAGCAGACGAGAGAGATTGTGGGAGTGA 944
Qy      1109  ACTGATGCCAAAGATGGAATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1168
Db      945  ACCTGAAGCAAGAGATGGAAGGGGCAATGAAACAAATGGTGGGAATGGTGGTGAAG 1004
Qy      1169  TAGAACAGTGAGAGAACCTTAGAGTTGTAGTTTCAGACCAACAGTGAATTCATTCCTTGA 1228
Db      1005  CAGACAGTGAGAGACCCGAGATCGTAGTCAGACACACGAGTGAATTTGACATTTCTTGA 1064
Qy      1229  TGATGCTATAGGTGAGAAAATACGGGAGAAAGTAGTGAAGGGCAATCCAAATCCAAG 1288
Db      1065  CGACGGTTACAGATGAGAAAATACGGCCAGAAAGTCGTTAAGGGAAACCCCAATCCAAG 1124
Qy      1289  GAGTTACTACAGTGACACACCCAGGATGTCAGTGAAGGACACGTCGAAAGAGCCTC 1348
Db      1125  AAGCTACTACAGTGACACACCCATCGTTGTCAGTGAAGGAACATGTGGAGAGAGATC 1184
Qy      1349  ACATGACCTAAGGCTGTGATCAACCTTATGAGGGAAGCAACCATGATGTTCTCTGC 1408
Db      1185  ACAGACATGAGAGCAGTAATCACACCTAGAAAGGGAACACCAACCATGATGTTCTCTGC 1244
Qy      1409  AGCCCGTGGCAGTGG 1423
Db      1245  AGCTCGTGGTAGCGG 1259

```

```

RESULT 13
AX651760  AX651760 1539 bp DNA linear PAT 22-MAR-2003
LOCUS
DEFINITION Sequence 606 from Patent WO03000898.
ACCESSION AX651760
VERSION AX651760.1 GI:29154578
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Chang, H. S., Chen, W., Cooper, B., Glazebrook, J., Goff, S. A., Hou, Y. M.,
Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 606 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
Location/Qualifiers
1..1539
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 506 a 355 c 309 g 369 t
ORIGIN
Query Match 16.0%; Score 308.2; DB 6; Length 1539;
Best Local Similarity 67.1%; Pred. No. 7.7e-64;
Matches 493; Conservative 0; Mismatches 218; Indels 24; Gaps 3;

```

```

Qy      710  AGATGATGGGTACAAATGGAGGAATATGGCCAAACAAAGTGAAGGAGTGAATCC 769
Db      528  AGAGATGGTTACAAATGGAGGAATATGGCCAAACAAAGTGAAGGAGTGAATCC 587

```

QY	770	AGAAGTTTATTACAAATGACATACCCCAATTCGCTTCAAGAGAAAGGTTGAGAGGTC	829
Db	588	TCGGAGTTACTATAAGTGTAATTCCTTCCCTAAATTTGTCCAAACGAAAGAAAGTGGAGAGATC	647
QY	830	TTTATAGTGCACAAATTTACTGAGATAGTTTATTAAGGGTACTCATACCAATCCTAAGCCTCA	889
Db	648	TTTGGAAAGTGCAGATCACAGAGATTGTGTATAAGGAAGCCACACCAATCCTAAACCTCA	707
QY	890	AAATACTAGGAGAAATCATCAAACTCCTCTTCTTTGCAATCCCTCATTCAAAATTCAT	949
Db	708	GTCTACTAGAAGATCTTCTTCGTCTTCTTCGACTTTTTCATTCAGCTGTGTACATGCCAG	767
QY	950	CAGAACTGAAATCC-----CAGATCAATCTTATGCCACACATGGAAGTGGACA	997
Db	768	TTTGGATCATATCGTCAAGCTTCTTCTGATCAGCTTAATTCCTAATAAGTCTTTCATCA	827
QY	998	AATGAGTTCAGCTGCCACCCAGAA-----AACTCATCAATATCAATTTGAGATGA	1048
Db	828	GTCTGATTTCTTTGGGATGCAACAGAGGATAATACTACTTCTGATTTCTGTGTGACGA	887
QY	1049	TCATTTTGGCAGAGTTCCCAAAAGTGTAAATCAGGAGGGGATCAATATGATGAAGATGA	1108
Db	888	TCAGTTTCGAACCAAGGCTC---ATCGATTGTTCAGCAGAGACGAAGAAGATTGTGGGATGA	944
QY	1109	ACCTGATGCCAAAAGATGGAAAAATTTGAAGGTGTAAAAATGAGGGTATGTTCAGCCCTCGAAG	1168
Db	945	ACCTGAGCAAAAGATGGAAGGGGACAAATGAACAAATGGTGGGAATGGTGTGGAG	1004
QY	1169	TAGAACAGTGCAGAGAACCTAGAGTTGTAGTTTGCAGACCAACCAAGTGCATTTATCTTTGA	1228
Db	1005	CAAGACAGTGCAGAGAGCCGAGAAATCGTAGTCAGACCAACAGAGTGTATTTGACATTTCTTGA	1064
QY	1229	TCATGGCTATAGTGCAGAAAAATACGGGCAGAAAAGTAGTGAGGGCAATCCAAATCCCAAG	1288
Db	1065	CGACGCTTACAGATGGGAAAAATACGGGCAGAAAAATCGTGTAAAGGAAACCCAAATCCCAAG	1124
QY	770	AGAAGTTTATTACAAATGACATACCCCAATTCGCTTCAAGAGAAAGGTTGAGAGGTC	829
Db	588	TCGGAGTTACTATAAGTGTAATTCCTTCCCTAAATTTGTCCAAACGAAAGAAAGTGGAGATC	647
QY	830	TTTATAGTGCACAAATTTACTGAGATAGTTTATTAAGGGTACTCATTAACCATCTCTAAGCCTCA	889
Db	648	TTTGGAAAGTGCAGATCACAGAGATTGTGTATAAGGAAGCCACACCAATCCTAAACCTCA	707
QY	890	AAATACTAGGAGAAATCATCAAACTCCTCTTCTTTGCAATCCCTCATTCAAAATTCAT	949
Db	708	GTCTACTAGAAGATCTTCTTCGTCTTCTTCGACTTTTTCATTCAGCTGTGTACATGCCAG	767
QY	950	CAGAACTGAAATCC-----CAGATCAATCTTATGCCACACATGGAAGTGGACA	997
Db	768	TTTGGATCATATCGTCAAGCTTCTTCTGATCAGCTTAATTCCTAATAAGTCTTTCATCA	827
QY	998	AATGAGTTCAGCTGCCACCCAGAA-----AACTCATCAATATCAATTTGAGATGA	1048
Db	828	GTCTGATTTCTTTGGGATGCAACAGAGGATAATACTACTTCTGATTTCTGTGTGACGA	887
QY	1049	TCATTTTGGCAGAGTTCCCAAAAGTGTAAATCAGGAGGGGATCAATATGATGAAGATGA	1108
Db	888	TCAGTTTCGAACCAAGGCTC---ATCGATTGTTCAGCAGAGACGAAGAAGATTGTGGGATGA	944
QY	1109	ACCTGATGCCAAAAGATGGAAAAATTTGAAGGTGTAAATGAGGGTATGTTCAGCCCTCGAAG	1168
Db	945	ACCTGAAACAAAGATGGAAGGGGACAAATGAAACAAATGGTGGGAATGGTGTGGAAG	1004
QY	1169	TAGAACAGTGCAGAGAACCTAGAGTTGTAGTTTGCAGACCAACCAAGTGCATTTATCTTTGA	1228
Db	1005	CAAGACAGTGCAGAGAGCCGAGAAATCGTAGTCAGACCAACGAGTGTATTTGACATTTCTTGA	1064
QY	1229	TGATGGCTATAGTGCAGAAAAATACGGGCAGAAAAATGAGGGGCAATCCAAATCCCAAG	1288
Db	1065	CGACGCTTACAGATGGGAAAAATACGGCCAGAAATCGTGTAAAGGAAACCCAAATCCAG	1124
QY	1289	GAGTTACTACAGTGCACACACCCAGGATGTCAGTAGGAGGACGCTGGGAAAGAGCCTC	1348
Db	1125	AAAGTACTACAAAGTGCACAAACCTCGTGTCCAGTAGGAGGAAACATGTGGAGAGAGCATC	1184
QY	1349	ACATGACCTAAGGGCTGTGATCACAACTTATGAGGAGAAAGCACACCAATGATTTCTTGC	1408
Db	1185	ACACGACATGAGCGCAGTAAATCACAACTTACGAAGGGGAAACACAAACCAAGTGTTCCTGC	1244
QY	1409	AGCCCGTGGCAGTGG 1423	
Db	1245	AGCTCGTGTAGCGG 1259	
RESULT 14			
AF509499			
LOCUS		1539 bp	mRNA linear
DEFINITION		Arabidopsis thaliana WRKY transcription factor 33 (WRKY33) mRNA,	
ACCESSION		complete cds.	
VERSION		AF509499	
KEYWORDS		AF509499.1 GI:21105638	
SOURCE		Arabidopsis thaliana (thale cress)	
ORGANISM		Arabidopsis thaliana	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
AUTHORS		1 (bases 1 to 1539)	
TITLE		Lippok, B. and Somasich, I.E.	
JOURNAL		Arabidopsis thaliana transcription factor WRKY33	
REFERENCE		Unpublished	
AUTHORS		2 (bases 1 to 1539)	
TITLE		Lippok, B. and Somasich, I.E.	
JOURNAL		Direct Submission	
FEATURES		Submitted (05-MAY-2002) Biochemistry, Max-Planck-Institut fur Zuchtungsforschung, Carl-von-Linne-Weg 10, Koeln, NRW D-50829, Germany	
		Location/Qualifiers	

Qy 1289 GAGTTACTACAGTGCACACACCCAGGATGTCCAGTGGAGGACACGTGGAAGAGCGCTC 1348
Db 1125 AAGCTACTACAGTGCACACACCCAGGATGTCCAGTGGAGGACACGTGGAAGAGCGCTC 1184
Qy 1349 ACATGACCTAAGGCTGTGATCAACAATTATGAGGGAAGACACACCATGATGTTCTTGC 1408
Db 1185 ACACGACATGAGAGCAGTAATCAACCTACGAGGGAAGACACACCATGATGTTCTTGC 1244
Qy 1409 AGCCCGTGGCAGTGG 1423
Db 1245 AGCTCGTGGTAGCGG 1259

RESULT 15
AX654743
LOCUS AX654743 1668 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 4613 from Patent WO0300898.
ACCESSION AX654743
VERSION AX654743.1 GI:29157557
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 0300898-A 4613 03-JAN-2003;
SYNGENTA PARTICIPATIONS AG (CH)
FEATURES
source
1..1668
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"
BASE COUNT 326 a 554 c 574 g 212 t 2 others
ORIGIN

Query Match 14.4%; Score 277.2; DB 6; Length 1668;
Best Local Similarity 54.8%; Pred. No. 2.5e-56;
Matches 696; Conservative 1; Mismatches 534; Indels 39; Gaps 6;

Qy 174 GGCTCTGGTGTCCCAATTCAGTCCACACACACCTCTCTGCGCTCTCTCTCTCTCTCTCT 233
Db 115 GCGGAGGTGTCTCAGGTACAGGCAATGACCCGCTCTCTCTCTCTCTCTCTCTCTCTCT 174
Qy 234 CCCATTTCT 293
Db 175 CCGGTGTGCGGTCT 234
Qy 294 GACTCGCGGTCT 353
Db 235 GGCTCGCGGTCT 291
Qy 354 GTTCTCTCAGCTTCAATTCGAAGCAGTTCAGGGGGATCAGCAATTTGTCAAGGAA 413
Db 292 GCGTCTGAGCATTCTGAGGCGGAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 351
Qy 414 GAAGACAAAAGCTTCTCAAAATTTCTCTTTTCCAAACCGGATCAGGACCTCTCTCTCTCT 473
Db 352 GGCAGGAGCAGCAGAGGAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 411
Qy 474 ACAGCAACATACAGTCTTCAAAATGTCAAGTTCAAAACACACAG---CCATGAGATTTT 530
Db 412 GAGGAGGCGGTGCGAAACAGCAGCTTTCCAGCCAGCGGTCTCCACCGGCGGCGGCGG 471
Qy 531 CAGGAGGCGGCGAAGCAGGATTAATTTTCTCTCAGAAAGGATATGATCAAACTGAA--- 587
Db 472 GAGGATACAGAGTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 531
Qy 588 ---AATCT 644

Search completed: January 20, 2004, 18:26:01
Job time : 7205 secs

Db 532 ATGACCGCGGTGCCAACCGCGCGAGCTTTCGGTGGCGCGCGCTTCCAGCGGACCTCGTCG 591
Qy 645 AGCAATGGGTTCCTCAATCCGATTTATGCAATATTACCCCCACAAATCTCAGAGCTTTAAAGTAGA 704
Db 592 GAGATGCGCCACAGGTGCGGGTGGCGGGGTACAGCCAGCCGCGAGTCCGAGAGGCGG 651
Qy 705 AGGTCAAGATGATGGGTACAAATTTGGAGGAAATATGGCCAAAAAACAAGTGAAGGAAAGTGA 764
Db 652 TCGTCGGACGAGCGGTACAACTGGCGCAAGTACGGGCAAGCAGGTGAAGGAGCGAG 711
Qy 765 AATCAAGAAGTTATTACAAATGACATACCCCAATTCCTTACAAAGAGAGAGTTGAG 824
Db 712 AACCCCGCAGCTTACTACAAATGCACTTCCGAACTGCCGACCAAGAGAGGTTGAG 771
Qy 825 AGTCTTTAGATGACAAATTTACTGAGATAGTTTATAAGGTACTCTCATCAACATCTCAAG 884
Db 772 CGGTCTCTGACGCGCAGATCACCGAGATCTGTGTACAGGGGACGCAACACACGCAAG 831
Qy 885 CCTCAAAATATCTAGGAGAACTCATCAAACTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 944
Db 832 CCGCAGAACACGCGCAGGAATCTCCGCTCGTGGCGCGCAGGTCTCTGC----- 880
Qy 945 TCCTCAGAACTGAAATCCAGATCAATCTTATGCCACACATGGAAGTGGCAAAATGGAT 1004
Db 881 -AGAGCGCGCGGACATGTCCGAGCATCTCTTCG-----GGGCGCATGTCCGCG 927
Qy 1005 TCAGCTGCCACCCCGCAAACTCATCAATATATGAGATGATGATGATGATGATGATGATGAT 1064
Db 928 ACAGCGCGCAGCGCCGAGAACTCTGCGGCTGCTTCGTTGACGACGAGATCAGAGTTGGC 987
Qy 1065 TCCCAAAAGT---GTAAATCAGAGGGGATGATATGATGAAGATGAACCTGATGCCAA 1121
Db 988 TCCCTCTCGCGCGGGAACGCGCGCGCAGAGTTCGACGACGACGAGCGGATTCGAAG 1047
Qy 1122 AGATGAAATTTGAAGTGAATAATGAGGTATGTAGCCCTTGGAGAGTGAACAGTGA 1181
Db 1048 AGATGAGGAAAGACGCTGACGCGGAGGGATCTCCATGGCTGGCAACCGGACGCTGCGG 1107
Qy 1182 GRACTTAGAGTTGATGTTAGACAAACAGTGCATGATGATGATGATGATGATGATGATGAT 1241
Db 1108 GAACCGAGGTGTCT 1167
Qy 1242 TGGAGAAATATACGCGGAGAAAGTGTGAAAGGGCAATCCAAATCCAAAGGATTTACTACAAG 1301
Db 1168 TGGAGAAAGTATCGCGCAGAGAGTGTGTGAGGGCAACCCGAAACCAAGGAGCTACTACAAG 1227
Qy 1302 TGACACACCCAGATGTCCAGTGAAGAGCAGCTGGAAGAGCCTTCAATGACCTAAG 1361
Db 1228 TGACACACCGCGGTGCT 1287
Qy 1362 GCTGTGATCAAACTTATGAGGGAAGCAACCATGATGTTCTCTGAGCGCGCTGGCAGT 1421
Db 1288 GCGTGTATCAACCTTACGAGGGGAAACACACACGATGCTGCTGCTGCTGCTGCTGCTGCT 1347
Qy 1422 GGCAGGCAAT 1431
Db 1348 GCGCGCTTT 1357

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 16:14:39 ; Search time 4188 Seconds
(without alignments)
11188.880 Million cell updates/sec

Title: US-09-890-811B-9

Perfect score: 1928

Sequence: 1 gcacgagtctatgcgcctct.....tcaaaaaaaaaaaaaaaaaa 1928

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_plg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	610.2	31.6	673	9	AW133440
c 2	603.4	31.3	662	13	BU550574
c 3	597.8	31.0	757	10	BE820590
c 4	583.8	30.3	746	10	BE658511

c	5	583.6	30.3	749	10	BE657634
	6	579	30.0	603	13	BQ295846
	7	568.8	29.5	572	13	BQ080806
	8	553.6	28.7	568	13	BU091659
	9	528.4	27.4	531	12	BI975061
	10	515	26.7	585	12	BM519918
	11	505.4	25.9	507	12	BM143621
	12	499.4	25.2	509	13	BU578260
	13	485.4	25.2	772	14	CB894319
	14	484.4	25.1	570	14	CD042161
	15	471.8	24.5	518	12	BM187988
	16	462.4	24.0	577	12	BM568389
c	17	458.4	23.8	559	12	BG840033
c	18	453.2	23.5	657	12	BI967766
	19	449.4	23.3	451	13	BU762115
	20	434	22.5	435	10	BE803423
	21	422.4	21.9	424	12	BM890972
	22	413.4	21.4	635	13	BU645573
	23	411	21.3	532	12	BM188894
	24	407.4	21.1	440	10	BG654191
	25	405.8	21.0	480	12	BM520930
	26	404.8	21.0	408	13	BU764331
	27	404.4	21.0	406	10	BG508622
	28	395.2	20.5	667	14	CD007418
	29	386.2	20.0	652	14	CD008237
	30	384.8	20.0	408	13	BU764518
	31	381.8	19.8	646	14	CD008324
	32	381.6	19.8	650	14	CD008395
	33	380.8	19.8	577	14	CA937181
	34	379	19.7	640	13	BU043758
	35	374.6	19.4	607	14	CA821401
	36	372.4	19.3	386	10	BG046426
	37	370	19.2	370	10	BE802905
	38	367.8	19.1	425	12	BM177021
	39	365.4	19.0	444	9	AW704231
	40	362.6	18.8	442	9	AW432500
	41	359.2	18.6	717	12	BI923269
	42	358.4	18.6	453	12	BM520345
	43	357.4	18.5	409	12	BI321844
	44	355	18.4	637	9	AI771129
	45	351.2	18.2	420	9	AU251455

ALIGNMENTS

RESULT 1
AW133440

LOCUS

DEFINITION

se19a06.y1 Gm-cl015 Glycine max cDNA clone

Gm-cl015-1643 5' similar to TR:Q40090 Q40090 SPPI

sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AW133440 673 bp mRNA linear EST 02-DEC-2001
se19a06.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl015-1643 5' similar to TR:Q40090 Q40090 SPPI PROTEIN. ; mRNA

AW133440 GI:6135047

AW133440.1

EST.

Glycine max (soybean)

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Glycine.

1 (bases 1 to 673)

Shoemaker.R., Keim.P., Vodkin.L., Erpelting.J., Corvill.V., Khanna

.A., Bolla.B., Marra.M., Hillier.L., Kucaba.T., Martin.J., Beck.C.,

Wyllie.T., Underwood.K., Steptoe.M., Theising.B., Allen.M., Bowers

.Y., Person.B., Swaller.T., Gibbons.M., Pape.D., Harvey.N., Schurk

.R., Ritter.E., Kohn.S., Shin.T., Jackson.Y., Cardenas.M., McCann

.R., Waterston.R. and Wilson.R.

Public Soybean EST Project

Unpublished

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com

Insert Length: 1363 Std Error: 0.00

Seq primer: -40RP from Gibco

High quality sequence stop: 416.

Location/Qualifiers

1. .673

/organism="Glycine max"

/mol_type="mRNA"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-1643"

/tissue_type="Mature flowers, field grown plants"

/lab_host="XL10-Gold"

/clone_lib="Gm-cl015"

/note="vector: pBluescript II XR; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelnding."

BASE COUNT 234 a 145 c 150 g 143 t 1 others

ORIGIN

Query Match 31.68; Score 610.2; DB 9; Length 673;

Best Local Similarity 95.58; Pred. No. 2e-97;

Matches 638; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

QY 808 CAAAGAGAGAGGTTGAGAGGCTTTAGATGGACAAATTTACTGAGATAGTTTATAAGGGTA 867

DB 6 CAAAGAGAGGTTGAGAGGCTTTAGATGGACAAATTTACTGAGATAGTTTATAAGGGTA 65

QY 868 CTCATACCATCTTAAGCCTCAAAATCTAGGAGAAATCTATCAAACTCTCTCTCTTG 927

DB 66 CTCATACCATCTTAAGCCTCAAAATCTAGGAGAAATCTATCAAACTCTCTCTCTTG 125

QY 928 CAATCCCTCATTCATCAATCCATCAGAACTGAAATCCAGATCAATCTATGCCACACATG 987

DB 126 CAATCCCTCATTCATCAATCCATCAGAACTGAAATCCAGATCAATCTATGCCACACATG 185

QY 988 GAAGTGACAAATGGATTGAGTCCACCCAGACAAATCTATCAATATCAATTTGGAGATG 1047

DB 186 GAAGTGACAAATGGATTGAGTCCACCCAGACAAATCTATCAATATCAATTTGGAGATG 245

QY 1048 ATGATTTTGACAGAGTTCCCAAAGTGTAATCAGAGGGGATGAATATGATGAAGATG 1107

DB 246 ATGATTTTGACAGAGTTCCCAAAGTGTAATCAGAGGGGATGAATATGATGAAGATG 305

QY 1108 AACCTGATGCCAAAGATGGAATTCGAGTGAATATGAGGATGATCTCAGCCCTCGAA 1167

DB 306 AACCTGATGCCAAAGATGGAATTCGAGTGAATATGAGGATGATCTCAGCCCTCGAA 365

QY 1168 GTAGAACAGTGAGAGAACCTTAGATTGAGTTTCAGACAAACAGTGCATATGATATCTCTTG 1227

DB 366 GTAGAACAGTGAGAGAACCTTAGATTGAGTTTCAGACAAACAGTGCATATGATATCTCTTG 425

QY 1228 ATGATGCTATAGTGGAGAAATATACGGGCGAAGATGATGAAGGGCAATCCAAATCCAA 1287

DB 426 ATGATGCTATAGTGGAGAAATATACGGGCGAAGATGATGAAGGGCAATCCAAATCCAA 485

QY 1288 GGAGTTACTACAGTGACACACCCAGGATGTCCTCAGTGAGGAGACACGTCGGAAGAGCCT 1347

DB 486 GGAGTTACTACAGTGACACACCCAGATGTCCAGTGAGGGAACACGTCGGAAGAGCCT 545

QY 1348 CACATGACCTAAGGCTGTGATCACAACTATGAGGGAAGACACCACTGATGTTCTCTG 1407

DB 546 CACATGACCTAAGGCTGTGATCACAACTATGATGGAAGACACCACTGATGTTCTCTG 605

QY 1408 CAGCCCGTGGCAGTGGCAG-CCATTTCTGTGAACAGACCAATGCCAAACAAATGTTCAAAC 1466

DB 606 CCACCCCGTGGCAGTGGCAGCCCATTTCTGTACAGACCAATGCCACCAATGTTCAAAC 665

QY 1467 CACACCAA 1474

DB 666 CACACCA 673

RESULT 2

BU550574/c

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU550574

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 662)

REFERENCE

the cDNA clones of the reracked Gm-r1088 library were then sequenced at the 3' end. The unigene selection and 3' sequencing was funded by NSF Plant Genome project #9872565 (http://soybeanomics.cropsci.uiuc.edu/) as part of creation of a low redundancy soybean cDNA set. The source cDNA libraries were constructed by the laboratories of Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa State University, and Paul Keim, Northern Arizona University as part of the Public EST project, http://129.186.26.94/soybeanest.html. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota. http://web.ahc.umn.edu/biodata/nfsoc/. Reracking and 3' sequencing were conducted by services of the University of Illinois Keck Center for Comparative and Functional Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The corresponding 5' EST from each clone in the Gm-r1088 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'.

BASE COUNT	179 a	145 c	146 g	192 t
ORIGIN	Query Match Best Local Similarity 31.3%; Score 603.4; DB 13; Length 662; Matches 604; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1287	AGGAGTTACTACAAAGTGCACACCCAGGATGTCAGTGGAGGAGGACGTTGGAAAGGCC	1346	
DB	605	AGGAGTTACTACAAAGTGCACACCCAGGATGTCAGTGGAGGAGGACGTTGGAAAGGCC	546	
QY	1347	TCAATGACCTAAGGCTGTGATCAACTTATGAGGGAAGGACACCATGATGTCCT	1406	
DB	545	TCAATGACCTAAGGCTGTGATCAACTTATGAGGGAAGGACACCATGATGTCCT	486	
QY	1407	GCAGCCGCTGGCAGTGGCAGCACTTCTGTAACAGACCAATGCAACCAATGCTTCAAC	1466	
DB	485	GCAGCCGCTGGCAGTGGCAGCACTTCTGTAACAGACCAATGCAACCAATGCTTCAAC	426	
QY	1467	CACACCAACTGCAGCCACTTCCGTAAGGCTCTTGCAGTGTATCCACCAAGTGACAAT	1526	
DB	425	CACACCAACTGCAGCCACTTCCGTAAGGCTCTTGCAGTGTATCCACCAAGTGACAAT	366	
QY	1527	TCCCTTCAGAACCAAGATCAAGACCAACGAGAGGCAATCACCTTCACCTAGAG	1586	
DB	365	TCCCTTCAGAACCAAGATCAAGACCAACGAGAGGCAATCACCTTCACCTAGAG	306	
QY	1587	ATGCTACAAAGTCCAGGAGTTTGGATCTCAGGTTTGGGAATCCAAATGCAATCTTAC	1646	
DB	305	ATGCTACAAAGTCCAGGAGTTTGGATCTCAGGTTTGGGAATCCAAATGCAATCTTAC	246	
QY	1647	GTGAACCCAGCAGCAACTATCTGACAAATGTTTCTCTCCAGGACCAAGGAGGAGCTAG	1706	
DB	245	GTGAACCCAGCAGCAACTATCTGACAAATGTTTCTCTCCAGGACCAAGGAGGAGCTAG	186	
QY	1707	GATGACATGTTCTGATGCTCTACTATGCTGAAGGAATTTTTTTTCCCTTTTGGTA	1766	
DB	185	GATGACATGTTCTGATGCTCTACTATGCTGAAGGAATTTTTTTTCCCTTTTGGTA	126	
QY	1767	GCTATGGAAGTTGGAAATTTTGGAGTGGGGGACTAGGATTTATTTGGACAAATAGGTT	1826	
DB	125	GCTATGGAAGTTGGAAATTTTGGAGTGGGGGACTAGGATTTATTTGGACAAATAGGTT	66	
QY	1827	CCATTGGAATTTATGCAATTTTGGTTGTTTGTGTTAAATTTTATACAGCCACAGGAT	1886	
DB	65	CCATTGGAATTTATGCAATTTTGGTTGTTTGTGTTAAATTTTATACAGCCACAGGAT	6	
QY	1887	TGGTA	1891	
DB	5	TGGTA	1	

RESULT 3
BE820590/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE820590 757 bp mRNA linear EST 24-MAY-2001
GM700012B20812 Gm-r1070 Glycine max cDNA clone Gm-r1070-4704 3',
mRNA sequence.

BE820590.1 GI:10252824

BE820590

EST.

Glycine max (soybean)

Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

1 (bases 1 to 757)

Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
Erpelidging,J., Raph,C., Shoop,E., Fardinas,J., Liu,L. and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)

Unpublished

Other ESTs: AW396234 corresponding to Gm-cl016-5873 (5')

Contact: Vodkin, L.O., PI, A Functional Genomics Program for

Soybean (NSF 9872565)

Lewin, H. A., Director, Keck Center for Comparative and Functional

Genomics

University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA

Tel: (217) 244-6147

Fax: (217) 333-4582

Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome

systems.com web site: www.genomesystems.com

Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

Location/Qualifiers

1..757

/organism="Glycine max"

/mol_type="mRNA"

/db_xref="taxon:3847"

/clone="Gm-r1070-4704"

/clone_lib="Gm-r1070"

/note="The library Gm-r1070 is a sequence-driven, reracked
set of 9,216 clones selected from cDNA libraries from
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones from
the different libraries was used to select singletons, or
a representative of each contig, which were reracked to
form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by
the laboratory of Ernest Retzel, Center for Computational
Genomics and Bioinformatics, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
. Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and 3' sequencing by the
Keck Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html. Note: The
corresponding 5' EST from each clone in the Gm-r1070
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under
'OTHER EST'."

BASE COUNT
ORIGIN

229 a 154 c 147 g 200 t 27 others

Query Match

31.0%; Score 597.8; DB 10; Length 757;

Best Local Similarity 95.5%; Pred. No. 2.9e-95;
Matches 611; Conservative 0; Mismatches 27; Indels 2; Gaps 1;

QY 1277 TCCAAATCCAAAGAGTTACTACAGTGCCACACACCCAGGATGTCAGTAGGAGACGCT 1336
DB 757 TCCAAATCCAAAGAGTTACTACAGTGCCACACCCAGGATGTCAGTAGGAGACGCT 698
QY 1337 GGAAGAGCCTACATGACCTAGGCTGTGATCACAACCTATGAGGGAAGCACAACCA 1396
DB 697 NNNANGAGCCTACATGANNAGGCTGTGATCACAACCTATGAGGGAAGCANNNCA 638
QY 1397 TGATGTTCTTCGAGCCGCTGCGAGTGGCAGCCATTCTGTGAACAGACCAATGCCAACAA 1456
DB 637 TGATGTTCTTCGAGCCGCTGCGAGTGGCAGCCATTCTGTGAACAGACCAATGCCAACAA 578
QY 1457 TGCTTCAAAACACACCAACACTGTCAGCCACTTCGTAAGGCTCTGCGAGTGCATCCACCA 1516
DB 577 TGCTTCAAAACACACCAACACTGTCAGCCACTTCGTAAGGCTCTGCGAGTGCATCCACCA 518
QY 1517 AAGTGACAATTCCTTCAGAACCAAGATCACAAGCACCACCAAGGGCAATCACCCCTT 1576
DB 517 AAGTGACAATTCCTTCAGAACCAAGATCACAAGCACCACCAAGGGCAATCACCCCTT 458
QY 1577 CACCTTAGAGATGCTACAAAGTCAGGAAGTTTGGATTCTCAGGGTTTGGGAATCCAAT 1636
DB 457 CACCTTAGAGATGCTACAAAGTCAGGAAGTTTGGATTCTCAGGGTTTGGGAATCCAAT 398
QY 1637 GCAATCTTAGTGAAACAGCAGCAACTATCTGACAATGTTTCTCCTCCAGGACCAAGGA 1696
DB 397 GCAATCTTAGTGAAACAGCAGCAACTATCTGACAATGTTTCTCCTCCAGGACCAAGGA 338
QY 1697 GGAGCCTAGAGATGATGCTTCTGAGTCTCTACTATGCTGAGGAA--TTTTTTTTTT 1754
DB 337 GGAGCCTAGAGATGATGCTTCTGAGTCTCTACTATGCTGAGGAAATTTTTTTTTTT 278
QY 1755 CCCTTTTGGTAGTATGGAAGTTGGAATTTTGGAGTGGGGACTAGGATTTATGG 1814
DB 277 CCCTTTTGGTAGTATGGAAGTTGGAATTTTGGAGTGGGGACTAGGATTTATGG 218
QY 1815 ACAATAAGGTTCCATTGATTTATGCAATTTTGTGTTTGTGTTGTAATTTTATA 1874
DB 217 ACAATAAGGTTCCATTGATTTATGCAATTTTGTGTTTGTGTTGTAATTTTATA 158
QY 1875 CAGCCACAGGATTCGTATGATATATATAGTATTTTCAAAA 1914
DB 157 CAGCCACAGGATTCGTATGATATATATAGTATTTTCAAA 118

RESULT 4
BE658511/c
LOCUS GM70006A20H3 Gm-r1070 Glycine max cDNA clone Gm-r1070-2262 3',
DEFINITION mRNA sequence.
ACCESSION BE658511
VERSION BE658511.1 GI:9984403
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosid I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 746)
Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V.,
Expelling, J., Raph, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished
Other ESTs: AW133440 corresponding to Gm-cl015-1643 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

Location/Qualifiers
1..746
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1070-2262"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked
set of 9,216 clones selected from cDNA libraries from
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones from
the different libraries was used to select singletons, or
a representative of each contig, which were reracked to
form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by
the laboratory of Ernest Retzel, Center for Computational
Genomics and Bioinformatics, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/Index.html
. Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and 3' sequencing by the
Keck Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html. Note: The
corresponding 5' EST from each clone in the Gm-r1070
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under
'OTHER EST'."

BASE COUNT 216 a 151 c 154 g 198 t 27 others
ORIGIN
Query Match 30.3%; Score 583.8; DB 10; Length 746;
Best Local Similarity 95.1%; Pred. No. 8.5e-93;
Matches 588; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY 1297 ACAAGTGCACACACCAGGATGTCAGTAGGAGGACGTCGAGGAGAGGCTCATGAGCC 1356
DB 746 ACNAGTGCACANNCCAGGATTTCCAGTGNNNAGCACGTTNNAGANNNNACANNANN 687
QY 1357 TAAGGGCTGTGATCACAACCTTATGAGGGAAGCACAACCATGATGTTCTCGAGCCGCTG 1416
DB 686 NAAGGGCTGTGATCACAACCTTATGAGGGAAGCACAACCATGATGTTNNNGCAGCCGCTG 627
QY 1417 GCAGTGCAGCCATTCTGTGAACAGACCATGCAACCATGCTTCAAAACACCAACCA 1476
DB 626 GCAGTGCAGCCATTCTGTGAACAGACCATGCAACCATGCTTCAAAACACCAACCA 567
QY 1477 CTCAGCCACTTCGTAAGGCTCTTGCAGTGTATCCACCAAGTGAATTCCTCTTCCAG 1536
DB 566 CTCAGCCACTTCGTAAGGCTCTTGCAGTGTATCCACCAAGTGAATTCCTCTTCCAG 507
QY 1537 ACCAAAGATCACAAGCACCACCAAGAGGGCAATCACCCCTTCAACCTAGAGATGCTACAA 1596
DB 506 ACCAAAGATCACAAGCACCACCAAGAGGGCAATCACCCCTTCAACCTAGAGATGCTACAA 447
QY 1597 GTCAGGAAGTTTGGATTTCTCAGGTTTGGGATCCAAATGCAATCTTACGTGAACGAGC 1656
DB 446 GTCAGGAAGTTTGGATTTCTCAGGTTTGGGATCCAAATGCAATCTTACGTGAACGAGC 387

QY 1657 AGCACTATCTGCAATGTTTCTCTCCAGGACCAAGGAGGAGCCTAGAGATGACATGT 1716
 |||||
 Db AGCACTATCTGCAATGTTTCTCTCCAGGACCAAGGAGGAGCCTAGAGATGACATGT 327
 |||||
 QY 1717 TCCTTGAGTCTCTACTATGCTGAAGGAATTTTTTTTCCCTTTTGGTAGCTATGGAAG 1776
 |||||
 Db TCCTTGAGTCTCTACTATGCTGAAGGAATTTTTTTTCCCTTTTGGTAGCTATGGAAG 267
 |||||
 QY 1777 GTTGGAAATTTTGAAGTGGGGAGCTAGGATTTATTGGACAAATAAGTTTCCATTCGATT 1836
 |||||
 Db GTTGGAAATTTTGAAGTGGGGAGCTAGGATTTATTGGACAAATAAGTTTCCATTCGATT 207
 |||||
 QY 1837 TATTGCATTTTGGTTTGTGTTGTAATTTTATACGCCACAGGATGTTGATAGTA 1896
 |||||
 Db TATTGCATTTTGGTTTGTGTTGTAATTTTATACGCCACAGGATGTTGATAGTA 147
 |||||
 QY 1897 TATACTAGTATTCAAA 1914
 |||||
 Db TATACTAGTATTCAAA 129
 |||||

RESULT 5

BE657634/c 749 bp mRNA linear EST 24-MAY-2001
 LOCUS GM700002B20B3 Gm-r1070 Glycine max cDNA clone Gm-r1070-462 3', mRNA
 DEFINITION sequence.

ACCESSION

BE657634

VERSION

BE657634.1 GI:9983526

KEYWORDS

EST.

SOURCE

Glycine max (soybean)

ORGANISM

Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE

1 (bases 1 to 749)
 Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V.,
 Erpelidig, J., Raph, C., Shoop, E., Pardini, J., Liu, L. and Lewin, H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 Unpublished

TITLE

Other ESTs: AW100733 corresponding to Gm-cl008-616 (5')
 Contact: Vodkin, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics

JOURNAL

University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodkin@uiuc.edu

COMMENT

This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134. For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or info@genome
 systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

source

1..749
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-462"
 /clone_lib="Gm-r1070"
 /note="The library Gm-r1070 is a sequence-driven, reracked
 set of 9,216 clones selected from cDNA libraries from
 various tissues and stages of development of soybean that
 represent 2,639 sequences from immature cotyledons, 1,770
 from immature seed coats, 3,938 from flowers, and 869
 from young pods. The 5' ESTs of the source clones from
 the different libraries was used to select singletons, or
 a representative of each contig, which were reracked to
 form library Gm-r1070. The cDNA clones of the reracked
 Gm-r1070 library were then sequenced at the 3' end. The

contig analysis to select unique genes was performed by
 the laboratory of Ernest Retzel, Center for Computational
 Genomics and Bioinformatics, University of Minnesota,
 http://www.cbc.umn.edu/ResearchProjects/soybean/index.html
 . Reracking was performed by Genome Systems, St. Louis,
 http://www.genomesystems.com, and 3' sequencing by the
 Keck Center for Comparative and Functional Genomics,
 University of Illinois,
 http://www.life.uiuc.edu/biotech/keck.html. Note: The
 corresponding 5' EST from each clone in the Gm-r1070
 library is listed in the 'OTHER EST' field. The detailed
 information on the source library for each clone can also
 be obtained by referring to the Genome Systems clone ID of
 the original cDNA library that is also listed under
 'OTHER EST'."

BASE COUNT 187 a 171 c 141 g 223 t 27 others
 ORIGIN
 Query Match 30.3%; Score 583.6; DB 10; Length 749;
 Best Local Similarity 88.3%; Pred. No. 9.2e-93;
 Matches 657; Conservative 0; Mismatches 70; Indels 17; Gaps 3;
 QY 1167 AGTAGAACAGTCGAGAACCTAGAGTTGTAGTTTCAGACAAACAGTCACATTTGATATCCCT 1226
 |||||
 Db 749 AGTAGNNNAGTAGAGNNNNNTAGAGTTGTAGTTTCAGNNNNNAGTCAGATTTGATATNNNA 690
 |||||
 QY 1227 GATGATGGCTATAGTGTGAGAAAATAACGGGCGAGAAAGTAGTGAAGGGCAATCCAAATCCA 1286
 |||||
 Db 689 GATGATGGCTATAGTGTGAGAAAATAACGGGCGAGAAAGTAGTGAAGGGCAATCCAAATCCA 630
 |||||
 QY 1287 AGAGTTTACTACAAGTGCACACCCAGGATGTCAGTGAGGAAGCAGCTGTGAAAAGGCC 1346
 |||||
 Db 629 AGAGTTTACTACAAGTGTACACNNCAGGATGTCAGTGAGGAAGCATGTGAAAAGAGCT 570
 |||||
 QY 1347 TCACATGACCTAAGCGCTGTGATCAACTTATGAGGGAAGACACCAATGATGTTCCT 1406
 |||||
 Db 569 TCACATGNNNAGGCGCTGTGATCAACTATGAGGGAAGACACCAATGATGTTCCT 510
 |||||
 QY 1407 GCAGCCCGTGGCAGTGGCAGCCATCTGTGAACAGACCAATGCAACCAATGCTTCAAAAC 1466
 |||||
 Db 509 GCAGCTCGTGGCAGTGGCAGCCATCTGTGAACAGACCAATGCAACCAATGCTTCAAAAC 450
 |||||
 QY 1467 CACACCAACATGTCAGCCCACTTCCTGTAAGGCTTGTGCCAGTGATCCACCAAGTGAACAT 1526
 |||||
 Db 449 CCTACCAACATGTCAGCCCACTTCCTGTAAGGCTTGTGCCAGTGATCCACCAAGTGAACAT 390
 |||||
 QY 1527 TCCTTCAGAACCAAGATCAGACCAACCAAGGAGGCAATCACCCTTCACCCCTAGAG 1586
 |||||
 Db 389 TCCTTCAGAACCAAGATCAGACCAACCAAGGAGGCAATCACCCTTCACCCCTAGAG 330
 |||||
 QY 1587 ATGCTACAAAGTCCAGAAAGTTTGGATTCTCAGGGTTTGGGAATCCAATGCAATCTTAC 1646
 |||||
 Db 329 ATGCTACAAAGTCCAGAAAGTTTGGATTCTCAGGGTTTGGGAATCCAATGCAATCTTAC 270
 |||||
 QY 1647 GTGAACACGAG-----CAACTATCTGCAATGTTTTCTCTCCAGGACCAAGGAGGAG 1700
 |||||
 Db 269 ATGAATCAGCAGCAACCAACTATCTGCAATGTTTTCTCTCCAGGACCAAGGAGGAG 210
 |||||
 QY 1701 CCTAGAGATGACATGTTCTCTGAGTCTCTACTATGCTGAGGAATTTTTTTCCTTT 1760
 |||||
 Db 209 CCTAGAGATGACATGTTCTCTGAGTCTCTACTATGCTGAGGAATTTTTTTCCTTT 159
 |||||
 QY 1761 TTGGTAGCTATGAAGTTTGGAAATTTTTCGNAAGTGGGGGACCTAGGATTTATTGGACAAAT 1820
 |||||
 Db 158 TTGGTAGCTATGAAGTTTGGAAATTTTTCGNAAGTGGGGGACCTAGGATTTATTGGACAAAT 99
 |||||
 QY 1821 AAGGTTCCATTCGATTTATTGCAATTTTGGTTTGTGTTTGTGTTAAATTTATACAGC-- 1878
 |||||
 Db 98 AAGGTTCCATTCGATTTATTGCAATTTTGGTTTGTGTTTGTGTTAAATTTATACAGCCA 39
 |||||
 QY 1879 CACAGATTTGGTATAGTATATATCT 1902
 |||||
 Db 38 CACAGGANNNGTATAGTATATANT 15
 |||||

RESULT 6
LOCUS BQ295846 603 bp mRNA linear EST 16-MAY-2002
DEFINITION sso27f11.y1 Gm-cl054 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-cl054-5637 5' similar to TR:Q40090 Q40090 SPFI PROTEIN. ;, mRNA sequence.
ACCESSION BQ295846
VERSION BQ295846.1 GI:20811368
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
REFERENCE 1 (bases 1 to 603)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished
COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: ccu@resgen.com web site: www.resgen.com
FEATURES Seg primer: -4ORF from Gibco
 High quality sequence stop: 421.
 Location/Qualifiers
 1..603
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl054-5637"
 /tissue_type="Leaf, 3 week old, greenhouse grown"
 /lab_host="DH10B"
 /clone_lib="Gm-cl054"
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The Harosoy NIL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht, 1995). The cDNA library was constructed from mRNA isolated from leaf tissue at various developmental stages of 3 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in cooperation with Dr. Paul Kelm's laboratory at Northern Arizona University."
 217 a 116 c 136 g 134 t
 BASE COUNT
 ORIGIN
 Query Match 30.0%; Score 579; DB 13; Length 603;
 Best Local Similarity 97.5%; Pred. No. 6.5e-92;
 Matches 588; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 755 GGGAGGTGAAAATCCAGAGTTATTACAAATGCACATACCCCAATTCCTTACAAAGAA 814
 |||||

Db 1 GGGAGGTGAAAATCCAGAGTTATTACAAATGCACATACCCCAATTCCTTACAAAGAA 60
 QY 815 GAAGGTTGAGAGTCTTTAGATGGACAAATTAAGTGGAGTAGTTTATAAGGGTACTCATAA 874
 Db 61 GAAGGTTGAGAGTCTTTAGATGGACAAATTAAGTGGAGTAGTTTATAAGGGTACTCATAA 120
 QY 875 CCATCCTAAGCCTCAAAATTAAGTGGAGAACTCATCAAACTCCTCTCTCTTCTTGGCAATCCC 934
 Db 121 CCATCCTAAGCCTCAAAATTAAGTGGAGAACTCATCAAACTCCTCTCTCTTCTTGGCAATCCC 180
 QY 935 TCATTCAAAATTCATCAGAACTGAATCCAGATCAATCCTATGCCACACATGGAAGTGG 994
 Db 181 TCATTCAAAATTCATCAGAACTGAATCCAGATCAATCCTATGCCACACATGGAAGTGG 240
 QY 995 ACAATATGGATTTCAGCTGCCACCCAGAAAACCTCATCAATATCAATTTGGAGATGATGATTT 1054
 Db 241 ACAATATGGATTTCAGCTGCCACCCAGAAAACCTCATCAATATCAATTTGGAGATGATGATTT 300
 QY 1055 TGAGCAGAGTTCCTCAAAAGTGAATTCAGAGGGGATGAATATGATGAAGATGAACCTGA 1114
 Db 301 TGAGCAGAGTTCCTCAAAAGTGAATTCAGAGGGGATGAATATGATGAAGATGAACCTGA 360
 QY 1115 TGCCTAAGATGGAATTCAGAGTGAATTCAGAGGGGATGAATATGATGAAGATGAACCTGA 1174
 Db 361 TGCCTAAGATGGAATTCAGAGTGAATTCAGAGGGGATGAATATGATGAAGATGAACCTGA 420
 QY 1175 AGTGAGAGAACCTTAGAGTTGTAGTTTCAGACCAACAGTGCATGATATCTTTCATGATGG 1234
 Db 421 AGTGAGAGAACCTTAGAGTTGTAGTTTCAGACCAACAGTGCATGATATCTTTCATGATGG 480
 QY 1235 CTATAGGTGGAGAAAATACGGGCAGAAAGTAGTGAAGGGCAATTCCTCAAGAGGAGTTA 1294
 Db 481 GTATAGATGAGAGAAAATACGGGCAGAAAGTAGTGAAGGGCAATTCCTCAAGAGGAGTTA 540
 QY 1295 CTCAAGTGCACACACCCAGAGTTCAGTGGAGAGACCTGGAGAGCCCTCAGATGA 1354
 Db 541 CTCAAGTGTACACACCCAGAGTTCAGTGGAGAGACCTGGAGAGCCCTCAGATGA 600
 QY 1355 CCT 1357
 Db 601 CCT 603
RESULT 7
LOCUS BQ080806 572 bp mRNA linear EST 04-APR-2002
DEFINITION ban10e08.y1 Gm-cl084 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-cl084-3663 5' similar to TR:Q40090 Q40090 SPFI PROTEIN. ;, mRNA sequence.
ACCESSION BQ080806
VERSION BQ080806.1 GI:19936302
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
REFERENCE 1 (bases 1 to 572)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished
COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com

Seq primer: -40RP from Gibco
High quality sequence stop: 421.

FEATURES

Location/Qualifiers
1..572
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl084-3663"
/tissue_type="Etiolated hypocotyls (Williams 82)"
/lab_host="DH10B"
/clone_lib="Gm-cl084"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed by M. Bhattacharya
from mRNA isolated from etiolated hypocotyls from the
cultivar Williams 82. Tissue was inoculated with
Phytophthora sojae race 1 and tissues were harvested 2 and
4 hours following infection. The library is the pool of
these two time points. Complementary DNA was synthesized
from mRNA using a primer consisting of a poly(dT) sequence
with a XhoI restriction site. EcoRI adapters were ligated
to the blunt-ended cDNA fragments followed by XhoI
digestion. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
DH10B host cells (Gibco BRL). This library was constructed
by M. Bhattacharya in the laboratory of Dr. Randy
Shoemaker at Iowa State University."
204 a 112 c 120 g 136 t

BASE COUNT 204 a 112 c 120 g 136 t

ORIGIN

Query Match 29.5%; Score 568.8; DB 13; Length 572;
Best Local Similarity 99.7%; Pred. No. 4.1e-90;
Matches 570; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 607 GTTTTTCCTGAGATTGCTGTCACAACTAACCATGACATGCTTCAATCCGATT 666
DB 1 GTTTTTCCTGAGATTGCTGTCACAACTAACCATGACATGCTTCAATCCGATT 60

QY 667 ATGCGAATTAACCCCAATCTCAGACTTTAAGTAGAAGGTGAGTGGTACAAT 726
DB 61 ATGCGAATTAACCCCAATCTCAGACTTTAAGTAGAAGGTGAGTGGTACAAT 120

QY 727 GAGGAAATATGCCAAAACCAAGTGAAGGAAAGTGAATTCAGAGTTATTAACAAT 786
DB 121 GAGGAAATATGTTCAAAAACCAAGTGAAGGAAAGTGAATTCAGAGTTATTAACAAT 180

QY 787 GCACATACCCCAATTCGCTTACAAAGAAAGGTTCAGAGTCTTTAGATGGCAATTA 846
DB 181 GCACATACCCCAATTCGCTTACAAAGAAAGGTTCAGAGTCTTTAGATGGCAATTA 240

QY 847 CTGAGATAGTTTATAAGGGTACTCATTAACCATCTTAAGCTCAAAATCTAGGAGAACT 906
DB 241 CTGAGATAGTTTATAAGGGTACTCATTAACCATCTTAAGCTCAAAATCTAGGAGAACT 300

QY 907 CATCAAACTCCTCTTCTTTGCAATCCCTCATTAATTCATCAGAACTGAAATCCCG 966
DB 301 CATCAAACTCCTCTTCTTTGCAATCCCTCATTAATTCATCAGAACTGAAATCCCG 360

QY 967 ATCAATCCTATGCACATGGAAGTGGACAAATGGATTGAGTCCACCCAGAAACT 1026
DB 361 ATCAATCCTATGCACATGGAAGTGGACAAATGGATTGAGTCCACCCAGAAACT 420

QY 1027 CATCAATATCAATTGGAGATGATGATTTTCAGCAGAGTTCCTCAAAAGTGTAAATCAGGAG 1086
DB 421 CATCAATATCAATTGGAGATGATGATTTTCAGCAGAGTTCCTCAAAAGTGTAAATCAGGAG 480

QY 1087 GGGATGAATATGATGAAGATGAACCTGATGCCAAAAGATGAAAATTTGAAGGTGAAAATG 1146

Db 481 GGGATGAATATGATGAAGATGAACCTGATGCCAAAAGATGAAAATTTGAAGGTGAAAATG 540

QY 1147 AGGGTATGTCAGCCCTGGAAGTAGAACAAGT 1178
|||||
Db 541 AGGGTATGTCAGCCCTGGAAGTAGAACAAGT 572
|||||

RESULT 8
BU091659 568 bp mRNA linear EST 29-AUG-2002
LOCUS Gm-cl054.y1 Gm-cl054 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl054-323 5' similar to TR:Q40090 Q40090 SPFI PROTEIN. ; mRNA
sequence.
ACCESSION BU091659
VERSION BU091659
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max

REFERENCE 1 (bases 1 to 568)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
TITLE Public Soybean EST Project
JOURNAL Unpublished
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 442.

FEATURES
Location/Qualifiers
1..568
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl054-323"
/tissue_type="Leaf, 3 week old, greenhouse grown"
/lab_host="DH10B"
/clone_lib="Gm-cl054"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from leaf tissue at various
developmental stages of 3 week old greenhouse grown
plants. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into DH10B host cells (GibcoBRL). This
library was constructed in cooperation with Dr. Paul
Keim's laboratory at Northern Arizona University."
198 a 121 c 130 g 119 t

BASE COUNT 198 a 121 c 130 g 119 t

Query Match 28.7%; Score 553.6; DB 13; Length 568;

Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
High quality sequence stop: 438.

FEATURES

source

959	Qy	AATCCCAGATCAATCCTATGCCACACATGGAAGTGGACAAATGGATTTCAGTGCACCCC	1018
61	Db	AATCCCAGATCAATCCTATGCCACACATGGAAGTGGACAAATGGATTTCAGTGCACCCC	120
1019	Qy	AGAAAACCTCATCAATATCAATTGGAGATGATGATTTTGACGACAGTTTCCCAAAGTGTAA	1078
121	Db	AGAAAACCTCATCAATATCAATTGGAGATGATGATTTTGACGACAGTTTCCCAAAGTGTAA	180
1079	Qy	ATCAGAGGGGATGAATATCATGAAGATGAACCTGATGCCAAAAGATGGAAAAATTGAAGG	1138
181	Db	ATCAGAGGGGATGAATATCATGAAGATGAACCTGATGCCAAAAGATGGAAAAATTGAAGG	240
1139	Qy	TGAAAATGAGGGTATGTTCAGCCCTCGGAAGTAGAACAGTGAAGAACTTAGAGTTGTAGT	1198
241	Db	TAAAATGAGGGTATGTTCAGCCCTCGGAAGTAGAACAGTGAAGAACTTAGAGTTGTAGT	300
1199	Qy	TCAGACAAACAGTGACATTGATTCCTTGATGATGCGCTATAGTGGAGAAAAATACGGGCA	1258
301	Db	TCAGACAAACAGTGACATTGATTCCTTGATGATGCGCTATAGTGGAGAAAAATACGGGCA	360
1259	Qy	GAAGTAGTGAAGGGCAATCCAAATCCAAAGGAGTTACTACAAGTGCACACACCCAGATG	1318
361	Db	GAAGTAGTGAAGGGCAATCCAAATCCAAAGAGTTACTACAAGTGCACACACCCAGATG	420
1319	Qy	TCAGTGCAGGAACGACGTGGAAAGAGCCTCACATGACCTAAGGGCTGTGATCACAACCTTA	1378
421	Db	TCAGTGCAGGAACGACGTGGAAAGAGCCTCACATGACCTAAGGGCTGTGATCACAACCTTA	480
1379	Qy	TGAGGGAAAAAGCAACCATGATGTTCTCGAGCCCGGTGGCAGTGGCAGCCATTCCTGTGAA	1438
481	Db	TGATGGAAGCAACAACCATGATGTTCTCGCAACCCGTGGCAGTGGCAGCCATTCCTGTGAA	540
1439	Qy	CAGACCAATGCCAAAACAATGCTCAAAAC	1466
541	Db	CAGACCAATGCCAAAACAATGCTCAAAAC	568

RESULT	9
BI975061	
LOCUS	
DEFINITION	BI975061 531 bp mRNA linear EST 30-NOV-2001 sai79508.y1 Gm-cl068 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl068-4552 5' similar to TR:Q9SXP5 Q9SXP5 TRANSCRIPTION FACTOR NTRWXY2. ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE	JOURNAL	COMMENT
1. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 2.
2. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1.
3. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2.
4. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3.
5. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4.
6. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5.
7. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5, 6.
8. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5, 6, 7.
9. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5, 6, 7, 8.
10. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5, 6, 7, 8, 9.

```

RESULT 10
BM519918          585 bp      mRNA      linear      EST 15-FEB-2002
LOCUS             sak86c03.y1 Gm-cl057 Glycine max cDNA clone SOYBEAN CLONE ID:
DEFINITION        Gm-cl057-2430 5' similar to TR:Q40090 Q40090 SPFI PROTEIN. ;, mRNA
sequence.
ACCESSION         BM519918
VERSION           BM519918.1 GI:18691070
KEYWORDS          Glycine max (soybean)
SOURCE            Glycine max
ORGANISM          Glycine max
                  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
                  ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
                  Glycine.
REFERENCE         1 (bases 1 to 585)
AUTHORS           Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
                  ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
                  Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
                  ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
                  ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
                  ,R., Waterston,R. and Wilson,R.
TITLE             Public Soybean EST Project
JOURNAL           Unpublished
COMMENT           Contact: Shoemaker R/Public Soybean EST Project
                  Washington University School of Medicine
                  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
                  Tel: 314 286 1800
                  Fax: 314 286 1810
                  Email: est@watson.wustl.edu
                  This clone is available through: ResGen, Invitrogen Corp. 2130
                  South Memorial Parkway Huntville, AL 35801 For further information
                  call: (800)-533-4363 or contact: ccu@resgen.com web site:
                  www.resgen.com
                  Seq primer: -40RP from Gibco
                  High quality sequence stop: 425.
FEATURES          source
                  1..585
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-cl057-2430"
                     /tissue_type="Degenerating cotyledons, 2 week old
                     seedling"
                     /lab_host="DH10B"
                     /clone_lib="Gm-cl057"
                     /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from degenerating cotyledons of 2 week old seedlings from
                     PI468916. Complementary DNA was synthesized from mRNA
                     using a primer consisting of a poly(dT) sequence with a
                     XhoI restriction site. EcoRI adapters were ligated to the
                     blunt-ended cDNA fragments followed by XhoI digestion.
                     The cDNA fragments were directionally cloned into the
                     EcoRI-XhoI restriction site of the pBluescript vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (GibcoBRL). This library was constructed in the
                     laboratory of Dr. Randy Shoemaker."
BASE COUNT        189 a 137 c 130 g 129 t
ORIGIN
Query Match      26.7%; Score 515; DB 12; Length 585;
Best Local Similarity 93.8%; Pred. No. 1.2e-80;
Matches 549; Conservative 0; Mismatches 30; Indels 6; Gaps 1;

Qy 1173 ACAGTGAGAGAACCTAGAGTTGTAGTTTCAGACACACAGTGCATTTGATATCTTGTATGAT 1232
      |||||
Db 1 ACAGTAGAGAACTAGAGTTGTAGTTTCAGACACACAGTGCATTTGATATCTTGTATGAT 60
      |||||

Qy 1233 GGCTATAGTGGAGAAATACGGGCAGAAAGTAGTGAAGGCGCAATCCAAATCCAAAGAGT 1292
      |||||

```

FEATURES

```

source
1. .507
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl072-2949"
/tissue_type="seedlings induced for symptoms of SDS
(Sudden Death Syndrome) disease"
/dev_stage="2-3 weeks old"
/lab_host="DH10B"
/clone_lib="Gm-cl072"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from 2-3 week old seedlings that were induced for symptoms
of SDS (Sudden Death Syndrome) disease by the
translocation of culture filtrate of Fusarium solani f.
sp. Glycines (Plant Cell Report 18:375-380). Cultivar PI
567374 is partially resistant to the disease SDS. Plant
tissue (expanded leaves, folded leaves, and new shoots)
were collected at 1, 6, 24, and 48 hrs. after inoculation
and their mRNA pooled equally for cDNA construction. The
library was prepared using the Stratagene pBluescript II
SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with an XhoI restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA insert is protected
from XhoI digestion via methylation during first strand
synthesis. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
E. coli ElectroMax DH10B host cells. Plants were inoculated
by Shuxian Li (Glen Hartman lab, University of Illinois).
Library was constructed by Steve Clough (Lila Vodkin lab,
University of Illinois)."
BASE COUNT      169 a      96 g      124 t
ORIGIN
Query Match      26.2%; Score 505.4; DB 12; Length 507;
Best Local Similarity 99.8%; Pred. No. 6.2e-79;
Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 306 CTCCTTAACCTCTCCAACTTCGCCATCTCCAACTCGAGCATTTGTGCTCAGAGC 365
DB 1 CTCCTTAACCTCTCCAACTTCGCCATCTCCAACTCGAGCATTTGTGCTCAGAGC 60
QY 366 TTCAATTGGAAGAGCAGTTTCAGGGGGGAATCAGCAAAATGTCAGGAGAGCAAAAGC 425
DB 61 TTCAATTGGAAGAGCAGTTTCAGGGGGGAATCAGCAAAATGTCAGGAGAGCAAAAGC 120
QY 426 TTCTCAAAATTTCTCTTTCCAAACCCGATCAGAGCTCTGCTTCATCCACGACACATAC 485
DB 121 TTCTCAAAATTTCTCTTTCCAAACCCGATCAGAGCTCTGCTTCATCCACGACACATAC 180
QY 486 CAGTCTTCAATGTTCACAGTTTCAACACACAGCATCGAGTGTTCAGGAGGCCACGAAA 545
DB 181 CAGTCTTCAATGTTCACAGTTTCAACACACAGCATCGAGTGTTCAGGAGGCCACGAAA 240
QY 546 CAGGATAATTTTCTCAGGAAAGGGTATGATGAAAACTGAAAACTCTTCTTCATCGAG 605
DB 241 CAGGATAATTTTCTCAGGAAAGGGTATGATGAAAACTGAAAACTCTTCTTCATCGAG 300
QY 606 AGTTTTTCCCTGAGATGCTAGTGTCCAACTAACCTAGCATGGTTCATCCGAT 665
DB 301 AGTTTTTCCCTGAGATGCTAGTGTCCAACTAACCTAGCATGGTTCATCCGAT 360
QY 666 TATGCAATTACCCCCCAATCTCAGACTTTTAAGTAGAAGTCAGATGATGGGTACAAAT 725
DB 361 TATGCAATTACCCCCCAATCTCAGACTTTTAAGTAGAAGTCAGATGATGGGTACAAAT 420
QY 726 TGGAGGAAATATGGCCAAAAAACAAGTGAAGGAAAGTGAAGTCCAAAGATTTATTACAAA 785
DB 421 TGGAGGAAATATGGTGAAGAAACAAGTGAAGGAAAGTGAAGTCCAAAGATTTATTACAAA 480

```

786 TGACATACCCCAATTGCTCTACAAAG 812

481 TGACATACCCCAATTGCTCTACAAAG 507

BU578260 509 bp mRNA linear EST 17-SBP-2002
sars0f02.y1 Gm-cl074 Glycine max cDNA clone SOYBEAN CLONE ID:
Gm-cl074-6003 5' similar to TR:Q40090 Q40090 SP1 PROTEIN. ;, mRNA
sequence.

BU578260
VERSION BU578260.1 GI:23057720
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max

RESULT 12
BU578260
LOCUS
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 509)
Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna
, A., Bolla, B., Marra, M., Hillier, L., Kucaba, F., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 405.
Location/Qualifiers
1. .509
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl074-6003"
/tissue_type="seedlings induced for HR (hypersensitive
response)"
/dev_stage="9-11 day old"
/lab_host="DH10B"
/clone_lib="Gm-cl074"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from 9-11 day old seedlings that were induced for HR
(hypersensitive response) by vacuum infiltrating plant
tissue with Pseudomonas syringae pv. glycinea carrying the
avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
and 53 hrs after inoculation and their mRNA pooled equally
for cDNA construction. The library was prepared using the
Stratagene pBluescript II SK(+) library construction kit.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly(dT) sequence with an XhoI restriction
site. EcoRI adaptors were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA insert is
protected from XhoI digestion via methylation during first
strand synthesis. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into E. coli ElectroMax DH10B host cells. Plant
care, inoculations, and library construction were

performed by Steve Clough (Lila Vodka lab, University of Illinois)."

BASE COUNT	183 a	101 c	118 g	107 t	
ORIGIN					
Query Match	25.9%	Score 499.4;	DB 13;	Length 509;	
Best Local Similarity	98.8%;	Pred. No. 7e-78;			
Matches	503;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;
QY	886	CTCAAAATCTAGGAGAACTCATCAAACTCCCTCTCTCTTGTGAATCCCTCATTTCAAAAT	945		
Db	1	CTCAAAATCTAGGAGAACTCATCAAACTCCCTCTCTCTTGTGAATCCCTCATTTCAAAAT	60		
QY	946	CCATCAGAACTGAAATCCCAAGATCAATCTATGACACATGGAAGTGGACAAATGGATT	1005		
Db	61	CCATCAGAACTGAAATCCCAAGATCAATCTATGACACATGGAAGTGGACAAATGGATT	120		
QY	1006	CAGTGCACCCAGAAAATCTATCAATATCAATTTGGAGATGATGATTTTGGACAGATT	1065		
Db	121	CAGTGCACCCAGAAAATCTATCAATATCAATTTGGAGATGATGATTTTGGACAGATT	180		
QY	1066	CCCAAAAGTCTAAATCAGGAGGGATGAATATGATGAAGATGAACCTGTGTCACAAAGAT	1125		
Db	181	CCCAAAAGTCTAAATCAGGAGGGATGAATATGATGAAGATGAACCTGTGTCACAAAGAT	240		
QY	1126	GGAAAATTGAAGTGAATAGGGTATGTGAGCCCTGGAAGTAGAACAAGTGGAGAAC	1185		
Db	241	GGAAAATTGAAGTGAATAGGGTATGTGAGCCCTGGAAGTAGAACAAGTGGAGAAC	300		
QY	1186	CTAGAGTTGAGTTTCAGAACACCAAGTACATTCCTTGTATGATGCTTATAGTGGGA	1245		
Db	301	CTAGAGTTGAGTTTCAGAACACCAAGTACATTCCTTGTATGATGCTTATAGTGGGA	360		
QY	1246	GAATAATCGGCGAGAAAGTAGTAGAGGGCAATCCAAATCCAAAGTATTACTACAGTGA	1305		
Db	361	GAATAATCGGCGAGAAAGTAGTAGAGGGCAATCCAAATCCAAAGTATTACTACAGTGA	420		
QY	1306	CACACCCAGATGTCCAGTGAAGGACGCTGGAAGAGGCTCACATGACCTTAAGGGCTG	1365		
Db	421	CACACCCAGATGTCCAGTGAAGGACGCTGGAAGAGGCTCACATGACCTTAAGGGCTG	480		
QY	1366	TGATCAAACTTATAGGGGAAAGCAAC	1394		
Db	481	TGATCAAACTTATAGGGGAAAGCAAC	509		
RESULT 13					
CB894319					
LOCUS	EST647111	HOGA Medicago truncatula cDNA clone HOGA-30F10, mRNA			
DEFINITION	772 bp	mRNA linear	EST 24-APR-2003		
ACCESSION	CB894319				
VERSION	CB894319.1	GI:30101488			
KEYWORDS	EST.				
SOURCE	Medicago truncatula (barrel medic)				
ORGANISM	Medicago truncatula				
REFERENCE	1 (bases 1 to 772)				
AUTHORS	Hahn M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,				
TITLE	Uterback, T., Cho, J. and Fraser, C.M.				
JOURNAL	ESTs from roots of Medicago truncatula treated with				
COMMENT	oligogalacturonides of DP 6-20				
	Unpublished				
	Contact: Michael G. Hahn				
	Complex Carbohydrate Research Center				
	University of Georgia				
	220 Riverbend Road, Athens, GA 30602-4712, USA				
	Tel: 706-542-4457				
	Fax: 706-542-4412				
	Email: hahn@ccrc.uga.edu				

TIGR sequence name: MTMDR29TK
More information is available at: www.medicago.org
Seq primer: SKmod (CTA GAA CTA gtg gat CC).

FEATURES

source

Location/Qualifiers					
1..772					
/organism="Medicago truncatula"					
/mol_type="mRNA"					
/cultivar="A17"					
/db_xref="taxon:3880"					
/clone="HOGA-30F10"					
/tissue_type="3 day old seedling roots"					
/dev_stage="24 hours after treatment in the dark at 26 C					
/with_0.5_mg_ml_oligogalacturonides (DP 6-20) in the					
/presence_of_100_u_g_ml_Gentamicin"					
/lab_host="XLOLR"					
/clone_lib="HOGA"					
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in SOLR cells."					
BASE COUNT	270 a	139 c	181 g	182 t	
ORIGIN					
Query Match	25.2%	Score 485.4;	DB 14;	Length 772;	
Best Local Similarity	81.4%;	Pred. No. 1.7e-75;			
Matches	616;	Conservative 0;	Mismatches 126;	Indels 15;	Gaps 4;
QY	684	CAATCTCAGACTTTAAGTAGAAGTCAGATGATGGGTACAAATTTGGAGGAAATATGGCCAA	743		
Db	16	CCAACCTCAAACTTTAAGCCGAGATCCGATGATGGATCAAAATTTGGCGAAATATGGACAA	75		
QY	744	AAACAAGTAGGAGGAAAGTGAATCCAGAAGTTATTACAATGACATACCCCAATGTC	803		
Db	76	AAGCAAGTTAAGGGAAGTGAATCCGAGAAGTTACTACAAATGACATACCCCTAACTGT	135		
QY	804	CCTACAAAGAAAGAGTTCAGAGGTCTTTAGATGGACAAATTAATCTGAGATAGTTTATAAG	863		
Db	136	CCTACAAAGAAAGAGTTCAGAGATCAATTTGAGGACAGTTACTGAGATAGTTTATAAG	195		
QY	864	GCTACTCATAACCATCTTAAGCCCTCAAAATCTAGGAGAAACTCATCAAACTCCTCTTCT	923		
Db	196	GCTACTCATAATCATCTTAAGCCCTCAATGTACGAGGAGAACTCGTCATCTCTTAAT	255		
QY	924	CTTGCAATCCTCATTTCAAAATTCATCAGAACTGAAATCCAGATCAATCTATGCCACA	983		
Db	256	GCTCTCGTGTGTTCTCTGTTAATCCAATCAATGAAATCCATGATCAATCATATGCATCA	315		
QY	984	CATGGAAGTGGACAAATGGATTTCAGCTGCCACCCAGAGAACTCATCAATATCAATTTGA	1043		
Db	316	CATGGTAATGGACAAATGGATTCTGCTGCTACTCCCGAGAAATCTTCAATATCAATAGGA	375		
QY	1044	---GATGATGATTTTTCAGCAGAGTTTC---CCAAAAGTGTAAATCAGGA---GGGGAT	1091		
Db	376	GGTGATGATGATTTTCAGCAGAGTTCTCACCAGAGGATAGATCAGGTGGTGGTGTAT	435		
QY	1092	GAATATGATGAAGATGAACCTGTATGTCACAAAGATGAAATTTGAAGGTGAAATAGGGGT	1151		
Db	436	GAATTTGATGAAGAGGAACCGGAGGCTTAAAGATGAAATTCGAAGTGAATATGAGGT	495		
QY	1152	ATGTCAGC---CCCTGGAAGTAGAACAGTGCAGAGAACCTAGAGTTCTAGTTCAGACAAAC	1208		
Db	496	ATATCAGCAACCTGCAAGTAGAATGTGTGAGAGAACTTAGAGTTGTAGTTTCAGACAAAC	555		
QY	1209	AGTGACATTGATATCTCTGATGATGCTATAGTGTGAGAGAAATACCGGCGAGAAAGTAGTG	1268		
Db	556	AGTGACATTGATATCTCTGATGATGCTATAGTGTGAGAGAAATACCGGCGAGAAAGTAGTG	615		
QY	1269	AAGGGCAATCCAAATCCAAAGGAGTTACTCAAGTGCACACACCCAGGATGTCCAGTGAGG	1328		
Db	616	AAGGGCAATCCAAATCCAAAGGAGTTACTCAAGTGCACACACCCAGGATGTCCAGTGAGG	675		

QY 1329 AAGCACTGAAAGAGCTCACATGACTAAGGCTGTGATCAACAATTATGAGGAAAG 1388
 Db 676 AAGCATGTTGAACGTGTTGCATGATTTAAGGCTGTGATCACTACTTATGAGGAAA 735
 QY 1389 CACAACCATGATGTTCTCGAGCCCGTGGCAGTGGCA 1425
 Db 736 CATAACCATGATGTCCTGCTGCTCGTGAAGCGAA 772

RESULT 14
 CD042161 570 bp mRNA linear EST 09-MAY-2003
 LOCUS psHB039x09f 301841 psHB: Infected hypocotyl soybean host. 48 hrs
 DEFINITION post infection Phytophthora sojae cDNA clone SHB039F09 5, mRNA
 sequence.
 ACCESSION CD042161
 VERSION CD042161.1 GI:30495754
 KEYWORDS EST.
 SOURCE Phytophthora sojae
 ORGANISM Phytophthora sojae
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.
 REFERENCE 1 (bases 1 to 570)
 AUTHORS Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
 TITLE USDA-IFAPS: Expression of Phytophthora sojae genes during infection and propagation
 JOURNAL unpublished
 COMMENT Contact: Tyler B
 Tyler lab
 VBI
 1980 Pratt Dr., Blacksburg, VA 24061, USA
 Tel: 540-231-7318
 Email: bmt Tyler@vt.edu
 PCR Primers
 FORWARD: BK reverse
 Plate: 039 row: F column: 09
 Seq primer: BK reverse
 High quality sequence stop: 570.
 Location/Qualifiers
 1..570
 /organism="Phytophthora sojae"
 /mol_type="mRNA"
 /db_xref="taxon:67593"
 /clone="SHB039F09"
 /tissue_type="infected host tissue"
 /cell_line="P6497"
 /dev_stage="48 hour post infection"
 /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs post infection"
 /note="vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; USDA-IFAPS: Expression of Phytophthora sojae genes during infection and propagation."
 BASE COUNT 152 a 161 c 98 g 159 t
 ORIGIN

Query Match 25.1%; Score 484.4; DB 14; Length 570;
 Best Local Similarity 92.1%; Pred. No. 2.9e-75;
 Matches 525; Conservative 0; Mismatches 36; Indels 9; Gaps 1;

QY 107 CTCGTACCTCTGCTTCTCCCTTGGACAAACAAGCCACACAG-----GGTGG 157
 Db 1 CTCGTACCTCTTGGCTTCTCCACAGACACAAACCACCATGTTGGTGGTGG 60
 QY 158 TTCTCTGAGAAACTGCTCTGCTGTGTTCCAAATCAAGTCCACACACACCTTCTCT 217
 Db 61 TTCTCTGAGAAACTGCTCTGCTGTGTTCCAAATCAAGTCCACACACACCTTCTCT 120
 QY 218 GCCTCTCTCCCTCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 277
 Db 121 GCCTCTCTCTCCCTCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
 QY 278 CCCTGCTGAGCTTCTTGACTCGCGGTTCTCTCTTAACCTCTTCCAACTTCTGCCATCTCC 337

Db 181 CCTGCTGAGCTTCTTGACTCACCTGTTCTCTTAACTTCTCCAACTTCTGCCATCTCC 240
 QY 338 AACAACCTGGAGCATTTGTTGCTCAGAGCTTCAATTGGAAGAGAGAGTTCAGGGGGGAATCA 397
 Db 241 AACAACCTGGAGCATTTGTTGCTCGGAGCTTCAATTGGAAGAGAGAGTTCAGGGGGGAATCA 300
 QY 398 GCAAATTTGCAAGGAAGAGACAAAGCTTCTCAAAATTTCTCTTCCAAACCCGATCAGG 457
 Db 301 GCGAATTTGCAAGGAAGAGACAAAGCTTCTCAAAATTTCTCTTCCAAACCCGATCAGG 360
 QY 458 ACCTCTGCTTTCATCCACAGCAACATACCATGCTTCCAAATGTCACAGTTCACAAACAACA 517
 Db 361 ACCTCTGCTTTCATCCACAGCAACATATCATGCTTTCAAATGTCACAGTTCACAAACAACA 420
 QY 518 GCCATGAGCTTTCAGGAGGCCACGAACAGAGATATTTTCTCTCAGGAAGAGGTATGAT 577
 Db 421 GCCATGAGCTTTCAGGAGGCCACGAACAGAGATATTTTCTCTCAGGAAGAGGTATGAT 480
 QY 578 GAAAACTGAAAACTCTTCTTCCATGACAGATTTTTCCTCCCTGAGATTTGCTAGTGTCCAAAC 637
 Db 481 GAAAACTGAAAACTCTTCTTCCATGACAGATTTTTCCTCCCTGAGATTTGCTAGTGTCCAAAC 540
 QY 638 TAACCATAGCAATGGGTTTCAATCCGATTA 667
 Db 541 TAACCATAGCAATGGGTTTCAATCCGATTA 570

RESULT 15
 BM187988 518 bp mRNA linear EST 11-DEC-2001
 LOCUS saj83d08.y1 Gm-cl074 Glycine max cDNA clone SOYBEAN CLONE ID:
 DEFINITION Gm-cl074-3232 5' similar to TR:O81639 O81639 ZINC FINGER PROTEIN.
 mRNA sequence.
 ACCESSION BM187988
 VERSION BM187988.1 GI:17518946
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 REFERENCE 1 (bases 1 to 518)
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,
 Y., Person,B., Swaller,T., Gibbons,M., Fape,D., Harvey,N., Schurk,
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Willson,R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estwaterston.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact: ccu@resgen.com web site:
 www.resgen.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 426.
 Location/Qualifiers
 1..518
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl074-3232"
 /tissue_type="seedlings induced for HR (hypersensitive response)"

/dev_stage="9-11 day old"
/lab_host="DH10B"
/clone_lib="Gm-cl074"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from 9-11 day old seedlings that were induced for HR
(hypersensitive response) by vacuum infiltrating plant
tissue with Pseudomonas syringae pv. Glycinea carrying the
avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
and 53 hrs after inoculation and their mRNA pooled equally
for cDNA construction. The library was prepared using the
Stratagene pBluescript II SK(+) library construction kit.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly(dT) sequence with an XhoI restriction
site. EcoRI adaptors were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA insert is
protected from XhoI digestion via methylation during first
strand synthesis. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into E. coli ElectroMax DH10B host cells. Plant
care, inoculations, and library construction were
performed by Steve Clough (Lila Vodkin lab, University of
Illinois)."

BASE COUNT 172 a 124 c 118 g 104 t
ORIGIN

Query Match 24.5%; Score 471.8; DB 12; Length 518;
Best Local Similarity 94.8%; Pred. No. 5e-73;
Matches 488; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 1150 GTATGTCAGCCCTGGAAAGTAGAACAGTGGAGAACCTAGAGTTGTAGTTTCAGACACCA 1209
Db 1 GTATGTCGCCCTGGAAAGTAGAACAGTGGAGAACCTAGAGTTGTAGTTTCAGACACCA 60
Qy 1210 GTGACATTGATATCCTTTGATGATGGCTATAGTGGAGAAAATACGGGCAGAAAGTAGTGA 1269
Db 61 GTGACATTGATATCCTTAGATGATGGGTATAGATGGAGAAAATACGGGCAGAAAGTAGTGA 120
Qy 1270 AGGGCATCCAAATCCAGAGGTACTACAGTGCACACCCAGGATGTCCAGTGAGGA 1329
Db 121 AGGCAATCAAAATCCAAAGAGTTACTACAAGTGTACACCCAGGATGTCCAGTGAGGA 180
Qy 1330 AGCAGTGGAAAGAGCCTCATGACCTAAGGGCTGTGATCAAACTTATGAGGGAAGC 1389
Db 181 AGCATGTGGAAGAGCTTCATGACCTAAGGGCTGTGATCAAACTTATGAGGGAAGC 240
Qy 1390 ACAACCATGATGTTCTTCGACCCCGTGGCAGTGGCAGCCATTCTGTGAACAGACCAATGC 1449
Db 241 ACAACCATGATGTTCTTCGACCTCGTGGCAGTGGCAGCCATTCTGTGAACAGACCAATGC 300
Qy 1450 CAACAAATGTTCAAAACACCAACCACTGAGCCACTTCGGTAAGSCTTTGCCAGTGA 1509
Db 301 CAACAAATGTTCAAAACCACTGAGCCACTTCGGTAAGSCTTTGCCAGTGA 360
Qy 1510 TCCACCAAGTGACAAATTCCTTCAGAAACAAAGATCAACAGCACCACAGAGGGCAAT 1569
Db 361 TCCACCAAGTGACAAATTCCTTCAGAAACAAAGATCTCAAGCACCACAGAGGGCAAT 420
Qy 1570 CACCCCTTACCCTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTCTCAGGGTTTGGGA 1629
Db 421 CACCCCTTACCCTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTCTCAGGGTTTGGGA 480
Qy 1630 ATCCAATGCAATCTTAGTGAACAGCAGCACTA 1664
Db 481 ATCCAATGCAATCTTAGTGAATCAGCAGCACTA 515

Search completed: January 20, 2004, 19:35:50
Job time : 4193 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:52 ; Search time 47 Seconds

(without alignments)
1941.868 Million cell updates/sec

Title: US-09-890-811B-10

Perfect score: 3050

Sequence: 1 MASSSGSLDTASANSFTNF.....FSSRTKEPRDDMFLESLIC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3050	100.0	575	22 AAE05092	Soybean SPFL-related
2	1605	52.6	560	23 ABG76909	Sunflower WRKY pol
3	1295	42.5	487	22 AAE05091	Rice SPFL-related
4	1139.5	37.4	512	21 AAG31843	Arabidopsis thaliana
5	1136.5	37.3	503	21 AAG31844	Arabidopsis thaliana
6	1117.5	36.6	488	22 AAE05090	Rice SPFL-related
7	904.5	29.7	619	22 AAE05093	Wheat SPFL-related
8	888.5	29.1	278	23 ABG76918	Wheat WRKY polypep.
9	884	29.0	309	21 AAG46847	Arabidopsis thaliana

10	861.5	28.2	514	24 ABP81179	Arabidopsis thaliana
11	858	28.1	577	23 ABG76916	Soybean WRKY polyp
12	852	27.9	687	23 AAU92971	Arabidopsis transc
13	848.5	27.8	513	24 ABP81191	Arabidopsis thaliana
14	818.5	26.8	279	21 AAG46848	Arabidopsis thaliana
15	742	24.3	568	23 AAU92967	Arabidopsis transc
16	702.5	23.0	200	21 AAG42218	Arabidopsis thaliana
17	702.5	23.0	216	21 AAG42217	Arabidopsis thaliana
18	702.5	23.0	216	21 AAG46849	Arabidopsis thaliana
19	625.5	20.5	406	23 ABG76910	Sunflower WRKY pol
20	619.5	20.3	413	23 ABG76911	Sunflower WRKY pol
21	618.5	20.3	583	23 ABG76922	Maize WRKY polypep
22	606	19.9	430	23 ABG76912	Sunflower WRKY pol
23	480	15.7	148	21 AAG42219	Arabidopsis thaliana
24	460.5	15.1	258	23 ABG76920	Maize WRKY polypep
25	406.5	13.3	189	21 AAG23979	Arabidopsis thaliana
26	400	13.1	163	21 AAG23980	Arabidopsis thaliana
27	363	11.9	68	23 AAO18478	A thaliana WRKY33
28	353	11.6	150	22 AAE05089	Corn SPFL-related
29	353	11.6	173	23 ABG76915	Soybean WRKY polyp
30	319.5	10.5	170	22 AAE05088	Corn SPFL-related
31	318	10.4	191	21 AAG17496	Arabidopsis thaliana
32	318	10.4	191	22 AAB67840	Amino acid sequenc
33	318	10.4	191	23 AAU92989	Arabidopsis transc
34	318	10.4	210	21 AAG17495	Arabidopsis thaliana
35	316.5	10.4	528	23 AAU93139	Arabidopsis transc
36	315.5	10.3	318	22 AAE01958	Arabidopsis thaliana
37	315.5	10.3	318	23 AAU93117	Arabidopsis transc
38	310.5	10.2	553	23 AAU93116	Arabidopsis transc
39	309.5	10.1	326	23 AAU92993	Arabidopsis transc
40	309.5	10.1	536	23 AAU92991	Arabidopsis transc
41	305.5	10.0	528	24 AAE30067	Arabidopsis thaliana
42	301	9.9	185	24 ABP96034	Nicotiana tabacum
43	301	9.9	415	21 AAG45532	Arabidopsis thaliana
44	301	9.9	545	21 AAG45531	Arabidopsis thaliana
45	300.5	9.9	406	21 AAG45533	Arabidopsis thaliana

ALIGNMENTS

RESULT 1
AAE05092
ID AAE05092 standard; Protein; 575 AA.
AC AAE05092;
XX
XX
DT 12-SEP-2001 (first entry)
XX
DE Soybean SPFL-related transcription factor #5.
XX
KW Soybean; SPFL-related transcription factor; transgenic plant;
XX genetical mapping; physical mapping; plant breeding.
XX
XX Glycine max.
XX
XX WO200149840-A2.
XX
PD 12-JUL-2001.
XX
XX 22-DEC-2000; 2000WO-US35310.
XX
XX 04-JAN-2000; 2000US-0174325.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX
XX Zhu Q, Famodu OO, Rafalski JA;
XX
XX WPI; 2001-441876/47.
XX
XX N-PSDB; AAD09829.
XX
XX Novel isolated SPFL-1 related transcription factor polypeptides and
XX polynucleotides useful for producing transgenic plants -

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140931.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Zhu Q, Famodu OO, Rafalski JA;
XX
DR WPI; 2001-441876/47.
DR N-PSDB; AAD09830.
XX
XX
PT Novel isolated SPF-1 related transcription factor polypeptides and
PT polynucleotides useful for producing transgenic plants -
XX
XX Claim 19; Page 56-58; 60pp; English.
XX
XX The present sequence is wheat SPF1-related transcription factor #6.
CC The SPF1-related transcription factor is useful for transforming a cell
CC by introducing SPF1-related transcription factor into a cell. It is also
CC useful for producing a transgenic plant by transforming a plant cell with
CC SPF1-related transcription factor and regenerating a plant from the
CC transformed plant cell. It is also useful to create transgenic plants in
CC which SPF1-related transcription factor DNA is present at higher or lower
CC levels than normal or in cell types or developmental stages in which they
CC are not normally found. The SPF1-related transcription factor DNA is
CC useful to prepare antibodies. It is also used as probes for genetically
CC and physically mapping the genes that they are a part of, and used as
CC markers for traits linked to these genes. Such information is useful in
CC plant breeding in order to develop lines with desired phenotype.
XX
XX Sequence 619 AA;
SQ

Query Match 29.7%; Score 904.5; DB 22; Length 619;
Best Local Similarity 39.2%; Pred. No. 6e-56;
Matches 228; Conservative 69; Mismatches 160; Indels 125; Gaps 21;

Qy 41 DNNKPPDGGLSERTGSGVPK-----FKSTPPPSLPLSPSPSSYFA 83
Db 9 DSPNPFSGDLPSAAGSPKPYADRRVAALAGAGARYKAMSPARLPISREPC-----LT 63
Qy 84 IPPCLSPAEILLDSVLLNSNLSPTTGAFVAGSFNWKSS-----SGGNQOIVKEEDK 137
Db 64 IPAGFSALLDSVLLTNKVEPSPTTGLSMAIMHKAHPDILSPDKSIRAHEDG 123
Qy 138 SFSNFSF-----QTRSGPPASSTATYQSSNVTVQTQOPWS 172
Db 124 GSRDFEFKPLHNSSSQSLAPAMSLDKKHEHSMQNSMNPSSS-----SSNNVWENRPPCS 178
Qy 173 FOEATKODNFSGK--GMM-KTENSMSQSFSEIASVQTNHNSGFSQDYGNYPPQSQT 228
Db 179 -RESSLTNVNSAPNPQVGMVGLTDNMPA-----EVGTSBPQQMN--SSDNAMQEPSEN 229
Qy 229 LSRRS-DDGYNWRKYQKQVKGSENPRSYKCTYPNCTPKKKKVERSLDGOITETVYKGTN 287
Db 230 VADKSADDGYNWRKYQKQVKGSENPRSYKCTHPNCEVKKLERAVDGLITEVYKGRH 289
Qy 288 NHPKPQNTRRNSNSSSLAIPHNSIRTEIPDQSYAT--HGSGQMDSAATPENS----- 339
Db 290 NHPKPQPNRLAGG---AVPSNQG--EERYDGAAGAAADKSSNALSNLANPNVSPGMVPE 343
Qy 340 -SISIGDDDFEQSQCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRVTRPRVVVQT 398
Db 344 VPVSVDSDDDIDAGGGR-PYPGDDATEEDLESKRRKMSAGIDAALMG-KPNRPREVVVQT 401
Qy 399 TSDIDILLDDGYRWRKYQKQVKGSENPRSYKCTHPGCPVKRKHVERASHDLRAVITYEG 458
Db 402 VSEVDILLDDGYRWRKYQKQVKGSENPRSYKCTSTGCPVKRKHVERASHDPKSVITYEG 461
Qy 459 KHNHDPVPAARGSGSHSVNRPMPNNAHNTNTAATSVRLLPVIHQSDNSL----- 507
Db 462 KHNHEVPAAR-NATHENSAPPMKN-----VHQINSMPSSIGGMMRAC 504
Qy 508 -----QNRQSAPEGSGSPFTLEMLQSPGSGFGFGNPNQS 544
Db 505 EARNFNSQYQSAEATDNLVSLDLGVGISP---NHSDATNQMS 543

RESULT 8
ABG76918
ID ABG76918 standard; Protein; 278 AA.
XX
AC ABG76918;
XX
DT 05-NOV-2002 (first entry)
XX
DE Wheat WRKY polypeptide #1.
XX
KW WRKY; plant; zinc-finger-type factor; WRKY; SAR; sunflower;
KW SA-dependent structure-activity analysis response; pathogen resistance;
KW maize; wheat; rice; soybean.
XX
OS Triticum aestivum.
XX
FN US2002076775-A1.
XX
PD 20-JUN-2002.
XX
PF 16-MAR-2001; 2001US-0810264.
XX
PR 17-MAR-2000; 2000US-190467P.
XX
PA (CRAN/) CRANE V C.
PA (FAMO/) FAMODU O.
PA (HUX/) HU X.
PA (JUGG/) LU G.
PA (ZHAN/) ZHANG L.
XX
XX Crane VC, Famodu O, Hu X, Lu G, Zhang L;
PI
DR WPI; 2002-582922/62.
DR N-PSDB; ABS59414.
XX
XX New isolated plant-specific zinc-finger-type factor polynucleotide,
PT useful for e.g. regulating the SA-dependent structure-activity analysis
PT response in a plant -
XX
XX Claim 9; Page 51-52; 66pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding a
CC plant-specific zinc-finger-type factor (WRKY) protein. The DNA and
CC protein are used to modulate the level of a WRKY protein in a plant and
CC to regulate the SA-dependent structure-activity analysis (SAR) response
CC in a plant. The sequences can be used to engineer plants to resist
CC pathogens such as viruses, bacteria, insects and fungi, and to survive
CC stresses. They may also be used as a molecular probe to track inheritance
CC of corresponding loci in genetic crosses and facilitate the plant
CC breeding process, to isolate, identify and genetically map WRKY and other
CC closely related disease resistance genes and to find genes and their
CC promoters that respond to a WRKY domain. This sequence represents a WRKY
CC polypeptide of the invention.
XX
XX Sequence 278 AA;
SQ

Query Match 29.1%; Score 888.5; DB 23; Length 278;
Best Local Similarity 52.4%; Pred. No. 2.7e-55;
Matches 184; Conservative 41; Mismatches 47; Indels 79; Gaps 11;

Qy 230 SRRSDDGYNWRKYQKQVKGSENPRSYKCTYPNCTPKKKKVERSLDGOITETVYKGTN 288
Db 2 NKQVEDGYNWRKYQKQVKGSENPRSYKCTYNNCSMKKKVERSLADGRITQIVYKGAHD 61
Qy 289 HPKPQNTRRNSNSSS-LAIPHNSIRTEIPDQSYATHGSGQMDSAATPENSISITGDD 347
Db 62 HPKPPSTRNSGCAAVIAEDHTN-----GSEHSGTTPENSIVTFGDDE 105
Qy 348 FEQSSQCKSGGDEYDEDEPAKRWKIEGENEGMS-APGS--RTVEPRVVVQTTSDIDI 404
Db 106 ADNGA-----EPETKRKRHEHGDNEGSGGTGACVAKPVRPLVUOTLSDIDI 152
Qy 405 LDDGYRWRKYQKQVKGSENPRSYKCTHPGCPVKRKHVERASHDLRAVITYEGKHNH 464

```
Db      153 LDDGFRWRKYGQVKGNPNRPSYKCTTVGCPVRKHVERASHDNRAVITTYEGRSHDV 212
Qy      465 PAARGSGSHSVNRPMPNNAASHNTAATSVRLPLVHQSDNSLQNRSQAPPGSQSPFTL 524
Db      213 PVGRGAGA---SRALPTSSSD-----SSVVVCPA-----AAGQAPYTL 248
Qy      525 EMLQSGSGFSGFGNPMQSYVNVQQLSDNVFSSRTKEPRDDMFLESLLC 575
Db      249 EMLANPAA-GRGYA-----AKDEPRDDMFVESLLC 278

RESULT 9
AAG46847
ID      AAG46847 standard; Protein; 309 AA.
XX
AC      AAG46847;
XX
DT      18-OCT-2000 (first entry)
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 58980.
XX
KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence.
XX
OS      Arabidopsis thaliana.
XX
PN      EF1033405-A2.
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      04-MAY-1999; 99US-0132407.
PR      05-MAY-1999; 99US-0132484.
PR      05-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      06-MAY-1999; 99US-0132487.
PR      07-MAY-1999; 99US-0132863.
PR      11-MAY-1999; 99US-0134256.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      14-MAY-1999; 99US-0134370.
PR      18-MAY-1999; 99US-0134768.
PR      19-MAY-1999; 99US-0134941.
PR      20-MAY-1999; 99US-0135124.
PR      21-MAY-1999; 99US-0135353.
PR      24-MAY-1999; 99US-0135629.
PR      25-MAY-1999; 99US-0136021.
PR      27-MAY-1999; 99US-0136392.
PR      28-MAY-1999; 99US-0136782.
PR      01-JUN-1999; 99US-0137222.
PR      03-JUN-1999; 99US-0137528.
PR      04-JUN-1999; 99US-0137502.
PR      07-JUN-1999; 99US-0137724.
PR      08-JUN-1999;
PR      10-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      14-JUN-1999; 99US-0138847.
PR      16-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      16-JUN-1999; 99US-0139453.
PR      17-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
PR      18-JUN-1999; 99US-0139750.
PR      18-JUN-1999; 99US-0139763.
PR      21-JUN-1999; 99US-0139817.
PR      22-JUN-1999; 99US-0139899.
PR      23-JUN-1999; 99US-0140353.
PR      23-JUN-1999; 99US-0140354.
PR      24-JUN-1999; 99US-0140695.
PR      28-JUN-1999; 99US-0140823.
PR      29-JUN-1999; 99US-0140991.
PR      30-JUN-1999; 99US-0141287.
PR      01-JUL-1999; 99US-0141842.
PR      01-JUL-1999; 99US-0142154.
PR      02-JUL-1999; 99US-0142055.
PR      06-JUL-1999; 99US-0142390.
PR      08-JUL-1999; 99US-0142803.
PR      09-JUL-1999; 99US-0142920.
PR      12-JUL-1999; 99US-0142977.
PR      13-JUL-1999; 99US-0143542.
PR      14-JUL-1999; 99US-0143624.
PR      15-JUL-1999; 99US-0144005.
PR      16-JUL-1999; 99US-0144085.
PR      16-JUL-1999; 99US-0144086.
PR      19-JUL-1999; 99US-0144325.
PR      19-JUL-1999; 99US-0144331.
PR      19-JUL-1999; 99US-0144332.
PR      19-JUL-1999; 99US-0144333.
PR      19-JUL-1999; 99US-0144334.
PR      19-JUL-1999; 99US-0144335.
PR      20-JUL-1999; 99US-0144352.
PR      20-JUL-1999; 99US-0144632.
PR      20-JUL-1999; 99US-0144884.
PR      21-JUL-1999; 99US-0144814.
PR      21-JUL-1999; 99US-0145086.
PR      21-JUL-1999; 99US-0145088.
PR      22-JUL-1999; 99US-0145085.
PR      22-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145089.
PR      22-JUL-1999; 99US-0145192.
PR      23-JUL-1999; 99US-0145145.
PR      23-JUL-1999; 99US-0145218.
PR      23-JUL-1999; 99US-0145224.
PR      26-JUL-1999; 99US-0145276.
PR      27-JUL-1999; 99US-0145913.
PR      27-JUL-1999; 99US-0145918.
PR      27-JUL-1999; 99US-0145919.
PR      28-JUL-1999; 99US-0145951.
PR      02-AUG-1999; 99US-0146386.
PR      02-AUG-1999; 99US-0146388.
PR      02-AUG-1999; 99US-0146389.
PR      03-AUG-1999; 99US-0147038.
PR      04-AUG-1999; 99US-0147204.
PR      04-AUG-1999; 99US-0147302.
PR      05-AUG-1999; 99US-0147192.
PR      05-AUG-1999; 99US-0147260.
PR      06-AUG-1999; 99US-0147303.
PR      06-AUG-1999; 99US-0147416.
```

PR	09-AUG-1999;	99US-0147493.	QY	58	VPKFKSTPPSLPLGPPPISSSYFAIPGLSPAELLDSVLNLSNIIPLSPPTTGAFVAQ	117
PR	09-AUG-1999;	99US-0147935.	Db	10	VPKFKTATP-----SPLPLSPSPYFTMPGLPADFLDSPLLFTSSNIIPLSPPTTGTFPAQ	64
PR	10-AUG-1999;	99US-0148171.	QY	118	SNWKSSSCG---NQIIVKEEDKSNFSPFQTRSGPPASSTATYQSSNVTVTQOPWSFQ	174
PR	12-AUG-1999;	99US-0148341.	Db	65	SLNY--NNGLLIDKNEIKYEDTT-----PP-----LFUPSMVT---QPLPOL	102
PR	13-AUG-1999;	99US-0148565.	QY	175	EATKQDNFSSGKMMKNTSSSMQSPSPSPEIASVQTNHNGFQSDYGNYPSPQSLTSSRRSD	234
PR	16-AUG-1999;	99US-0149358.	Db	103	DLFKSEIMSSNK-----	117
PR	17-AUG-1999;	99US-0149175.	QY	235	DGYNMRKYQKQKGVGSENPRSYKCTYPCNCTKKKVERSL-DQITEIVYKGHNHKPKQ	293
PR	18-AUG-1999;	99US-0149426.	Db	118	DGYNMRKYQKQKGVGSENPRSYKCTYPCNCTKKKVERSL-DQITEIVYKGHNHKPKQ	177
PR	20-AUG-1999;	99US-0149723.	QY	294	NTRNSNSSSLAIPHNSIRTEIPQSYATHGSGQMDSAATPENSISIGDDDFQSSQ	353
PR	20-AUG-1999;	99US-0149929.	Db	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	23-AUG-1999;	99US-0149902.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	23-AUG-1999;	99US-0150566.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	25-AUG-1999;	99US-0150566.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	26-AUG-1999;	99US-0150884.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	27-AUG-1999;	99US-0151066.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	27-AUG-1999;	99US-0151080.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	30-AUG-1999;	99US-0151303.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	31-AUG-1999;	99US-0151438.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	01-SEP-1999;	99US-0151930.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	07-SEP-1999;	99US-0152363.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	10-SEP-1999;	99US-0153070.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	13-SEP-1999;	99US-0153758.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	15-SEP-1999;	99US-0154018.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	16-SEP-1999;	99US-0154039.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	20-SEP-1999;	99US-0154779.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	22-SEP-1999;	99US-0155139.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	23-SEP-1999;	99US-0155486.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	24-SEP-1999;	99US-0155659.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	28-SEP-1999;	99US-0156458.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	29-SEP-1999;	99US-0156596.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	04-OCT-1999;	99US-0157117.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	06-OCT-1999;	99US-0157753.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	07-OCT-1999;	99US-0157865.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	08-OCT-1999;	99US-0158029.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	08-OCT-1999;	99US-0158232.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	12-OCT-1999;	99US-0158369.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	13-OCT-1999;	99US-0159293.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	13-OCT-1999;	99US-0159294.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	13-OCT-1999;	99US-0159295.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	14-OCT-1999;	99US-0159329.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	14-OCT-1999;	99US-0159330.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	14-OCT-1999;	99US-0159331.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	14-OCT-1999;	99US-0159637.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	14-OCT-1999;	99US-0159638.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	18-OCT-1999;	99US-0159584.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	21-OCT-1999;	99US-0160741.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	21-OCT-1999;	99US-0160767.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	21-OCT-1999;	99US-0160768.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	21-OCT-1999;	99US-0160770.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	21-OCT-1999;	99US-0160814.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	21-OCT-1999;	99US-0160815.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	22-OCT-1999;	99US-0160980.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	22-OCT-1999;	99US-0160981.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	22-OCT-1999;	99US-0160989.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	25-OCT-1999;	99US-0161404.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	25-OCT-1999;	99US-0161405.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	25-OCT-1999;	99US-0161406.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	26-OCT-1999;	99US-0161359.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	26-OCT-1999;	99US-0161360.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296
PR	26-OCT-1999;	99US-0161361.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	28-OCT-1999;	99US-0161920.	QY	354	KCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDILDDGYRWK	413
PR	28-OCT-1999;	99US-0161992.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKEPRVVVQTTSDIDILDDGYRWK	241
PR	28-OCT-1999;	99US-0161993.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	468
PR	29-OCT-1999;	99US-0162142.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITYEGKHNHDPAAAR	296

Query Match 29.0%; Score 884; DB 21; Length 309;
Best Local Similarity 47.7%; Pred. No. 6.5e-55;
Matches 198; Conservative 30; Mismatches 55; Indels 132; Gaps 13;

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145227.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Matches 184; Conservative 28; Mismatches 50; Indels 127; Gaps 12;	
Qy	84 IPGLSPAELLDSPVLLNSSNLPSPPTTGAFVAQSFNWKSSSG---NOQIVKEEDKFS 140
Db	1 MPFGLTPADFLDPLFTSSNLPSPPTTGTFPAQSLNY--NNGLLIDKNEIKYEDT-- 56
Qy	141 NFSFQTRSGPPASSTATYQSSNVTVTQQPWSFQEA TKQDNFSGGKGMKMTENSSSQSF 200
Db	57 -----PP-----LFLPSMT-----QPLQLDLFKSEIMSSNK----- 84
Qy	201 SPEIASVQTHNSGFSQDYGNYPPQSOTLSRRSDGYNWRKYGOKVKGSENPRSYKCT 260
Db	85 -----TSDDGYNWRKYGOKVKGSENPRSYPKCT 113
Qy	261 YPNCPTKKVERSIL-DQOITEIVVYKTHNHPKPNTRNNSNSSSLAIPHSNIRTEIPD 319
Db	114 YPNCLTKKVVETSLVKQOMIEIVVYKSHNHPKQSTKRSSTAIA---AHQNS----- 163
Qy	320 QSYATHSGQWDSAAATPENSISIGDDDFQSSQKCKSGGDEYDEBPDAAKWKIEGENE 379
Db	164 -----SNGD-----GKDIGEDETEAKRWKRE-EN- 186
Qy	380 GMSAPGSRVREPRVVVQTTSDIDLDGYNWRKYGOKVKGSENPRSYKCTHPGCPVR 439
Db	187 -----VKEPVVVQTTSDIDLDGYNWRKYGOKVKGSENPRSYKCTFTGCFVR 237
Qy	440 KHVERASHDLRAVITTYEGKHNDVPAAR 468
Db	238 KHVERAFQPKSVITTYEGKHQIPTPR 266
RESULT 15	
ID	AAU92967
XX	AAU92967 standard; Protein; 568 AA.
AC	AAU92967;
XX	
DT	02-JUL-2002 (first entry)
XX	
DE	Arabidopsis transcription factor #5.
XX	
KW	Agriculture; metabolic chemical; environmental stress; drought;
KW	microbial disease resistance; herbicide resistance; seed yield;
KW	fruit yield; growth rate; leaf senescence; flower senescence.
KW	plant; transcription factor; transgenic.
XX	
OS	Arabidopsis thaliana.
XX	
FN	WO200215675-A1.
XX	
PD	28-FEB-2002.
XX	
XX	22-AUG-2001; 2001WO-US26189.
XX	
PR	22-AUG-2000; 2000US-227439P.
PR	16-NOV-2000; 2000US-0713994.
PR	16-APR-2001; 2001US-0837944.
XX	(MEND-) MENDEL BIOTECHNOLOGY INC.
PA	(PILG/) PILGRIM M.
PA	(CREE/) CREELMAN R.
PA	(DUBE/) DUBELL A J.
PA	(HEAR/) HEARD J.
PA	(JIAN/) JIANG C.
PA	(KEDD/) KEDDIE J.
PA	(ADAM/) ADAM L.
PA	(RATC/) RATCLIFF O.
PA	(REUB/) REUBER J L.
PA	(RIEC/) RIECHMANN J L.
PA	(YUGG/) YU G.
PA	(PINE/) PINEDA O.
XX	
PI	Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddle J;

PI	Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX	
DR	WPI; 2002-292022/33.
DR	N-PSDB; ABR65153.
XX	
PT	An isolated or recombinant polynucleotide used to produce a transgenic plant -
XX	
PS	Claim 40; Page 89-91; 941pp; English.
XX	
CC	The invention relates to 1 of 232 isolated or recombinant polynucleotides encoding an Arabidopsis thaliana transcription factor, their variants, complements, fragments, or related polynucleotide with 31% to 95% sequence identity, where the plant possesses an altered trait as compared to a wild-type or reference plant, or the plant exhibits an altered phenotype as compared to a wild-type or reference plant, or the plant exhibits ectopic expression or altered expression of one or more genes associated with a plant trait as compared to a wild plant. Also included are a transgenic plant comprising the polynucleotides, a computer readable medium having stored sequence information, and identifying a homologue sequence from a database comprising a plurality of known plant sequences comprising inputting sequence information selected from one of 464 fully defined sequences given in the specification. The isolated or recombinant polynucleotide is used for producing a plant having a modified trait, the method comprising selecting a polynucleotide that encodes a polypeptide or an antisense nucleic acid, inserting the polynucleotide or antisense nucleic acid into an expression vector, introducing the vector into a plant or a cell of a plant to overexpress the polypeptide or antisense nucleic acid, thereby producing a modified plant, and selecting for a modified trait (e.g. increased production of agriculturally useful proteins or metabolic chemicals, pest tolerance, environmental stress response (e.g. drought), microbial disease resistance, herbicide resistance, seed and fruit yield, growth rate, leaf and flower senescence and many other traits listed in the specification). The present sequence is one of the 232 proteins which are A. thaliana transcription factors.
XX	
Seq	Sequence 568 AA;
Query Match 24.3%; Score 742; DB 23; Length 568;	
Best Local Similarity 37.3%; Pred. No. 2.1e-44;	
Matches 193; Conservative 55; Mismatches 119; Indels 150; Gaps 16;	
Qy	86 PGLSPAELLDSPVLLNSSNLPSPPTTGAFVAQSFNWKSSSGGQVQIVKEEDKFSFQ 145
Db	88 PGLSPATLLESPVFL--SNPLSPPTGKL-----SSVPSDKAKAELFDDITSLAQ 137
Qy	146 TRSGPPASSTATYQSSNVTVTQQPWSFQEA TKQDNFSGGKGMKMTENSSSQSFPEIA 205
Db	138 TISGGLDPT-----NIALEPDDSDYEERQ-----LGLG-----DSMACCAP--- 176
Qy	206 SVQTHNSGFSQDYGNYPPQSOTLSRRSDGYNWRKYGOKVKGSENPRSYKCTVNCNP 265
Db	177 -----ADGYNWRKYGOKVKGSEYPRSYKCTHPCE 209
Qy	266 TKKVERSILDGOITEIVVYKTHNHPKPNTRNNSNSSSLA----- 306
Db	210 AKKVERSREGHIEIITYGDHIHSPPNRKSIGSGSGTGQDMQMDATEYEGFAGTNE 269
Qy	307 IPHSNIRTEIPDQSYA-----THSGQWDSAAATPENSISIGDDDFQSSQKCKSG 358
Db	270 IEWTSVPVSAELEYGSHSGMQVQNGTHQFGYDAAA---DALYRDENEDDRTHSMVSL 325
Qy	359 GDEYDEBPDAAKWKIEG---ENEGSAPGSRVREPRVVVQTTSDIDLDGYNWRKYG 415
Db	326 TYDGEVESESKRKLEAYATETSG----STRASREPRVVVQTTSDIDLDGYNWRKYG 381
Qy	416 QKVKGNNPRSYKCTHPGCPVRKHVERASHDLRAVITTYEGKHNDVPAAR-----GS 470
Db	382 QKVKGNNPRSYKCTANGCTVTRHVERASDDFKSVLITYIGKHTVTPAARNSSHVGA 441
Qy	471 GS-----HSVNRMPNPNASNTHTAATSVRLLPVHIOSDNLQNRSQAPP 516

Db 442 GSSGTLQGLATQTHNHNHVPMPHSRSEGLATA-----NSSLDFDQSH--- 485
Qy 517 EGQSPFTLEMLQSPGSGFGSGFGNPMQSYVNOQQLSD 553
Db 486 -----LRHPTGFSV-----YIGQSELSLSD 503

Search completed: January 20, 2004, 16:19:24
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:52 ; Search time 21 Seconds
(without alignments)
1158.512 Million cell updates/sec

Title: US-09-890-811B-10
Perfect score: 3050
Sequence: 1 MASSGSLDTSASANSFTNF.....FSSRTKEPRDDMFLESLLC 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfile1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	5.3	1312	3	US-09-041-886-19
2	163	5.3	1312	4	US-09-648-281-2
3	154	5.0	2842	1	US-07-741-940-7
4	154	5.0	2842	1	US-08-289-548A-7
5	154	5.0	2842	1	US-08-452-654-7
6	154	5.0	2842	4	US-08-449-731-7
7	154	5.0	2843	1	US-07-741-940-2
8	154	5.0	2843	1	US-08-289-548A-2
9	154	5.0	2843	1	US-08-452-654-2
10	154	5.0	2843	1	US-08-452-655B-2
11	154	5.0	2843	1	US-08-452-655B-7
12	154	5.0	2843	2	US-08-370-235A-2
13	154	5.0	2843	3	US-08-450-582-2
14	154	5.0	2843	3	US-08-450-582-7
15	154	5.0	2843	4	US-08-449-731-2
16	154	5.0	2973	2	US-08-821-355A-7
17	154	5.0	2973	2	US-09-003-687A-7
18	154	5.0	2973	3	US-09-136-605-7
19	149.5	4.9	2137	4	US-09-134-001C-4463
20	147.5	4.8	951	4	US-09-125-635-8
21	147.5	4.8	1420	4	US-09-125-635-4
22	139	4.6	725	3	US-09-234-332-12
23	137.5	4.5	1404	4	US-08-801-308-1
24	135	4.4	1911	4	US-09-854-856-64
25	135	4.4	1939	4	US-09-854-856-48
26	135	4.4	1971	4	US-09-854-856-32
27	135	4.4	1999	4	US-09-854-856-16

28	135	4.4	2004	4	US-09-854-856-58	Sequence 58, Appl
29	135	4.4	2032	4	US-09-854-856-42	Sequence 42, Appl
30	135	4.4	2064	4	US-09-854-856-26	Sequence 26, Appl
31	135	4.4	2092	4	US-09-854-856-10	Sequence 10, Appl
32	135	4.4	2157	4	US-09-854-856-52	Sequence 52, Appl
33	135	4.4	2185	4	US-09-854-856-36	Sequence 36, Appl
34	135	4.4	2217	4	US-09-854-856-20	Sequence 20, Appl
35	135	4.4	2245	4	US-09-854-856-4	Sequence 4, Appl
36	134	4.4	3969	3	US-08-061-376-5	Sequence 5, Appl
37	132	4.3	467	4	US-09-134-001C-3235	Sequence 3235, Ap
38	132	4.3	2133	2	US-08-670-707A-37	Sequence 37, Appl
39	132	4.3	2133	3	US-09-037-601-37	Sequence 37, Appl
40	132	4.3	2133	4	US-09-315-179-37	Sequence 37, Appl
41	132	4.3	2133	4	US-09-523-656-30	Sequence 30, Appl
42	131.5	4.3	1402	4	US-09-125-635-12	Sequence 12, Appl
43	130.5	4.3	868	1	US-07-864-004B-6	Sequence 6, Appl
44	130.5	4.3	868	1	US-08-251-937A-6	Sequence 6, Appl
45	130.5	4.3	868	1	US-08-212-133A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-041-886-19
; Sequence 19, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Radzideh, Sharroz
; TITLE OF INVENTION: Propeptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-19

Query Match 5.3%; Score 163; DB 3; Length 1312;
Best Local Similarity 19.9%; Pred. No. 9.2e-05;
Matches 120; Conservative 76; Mismatches 260; Indels 146; Gaps 22;

QY	37	ASPLDNKKPQCGLSERTGSGVPKF--KSTPP-----PSLPL-----	71
DB	658	ATFPVARTPSGCTSSVSVGVPRLSPKTHRPSPQNSIGNTPSGFVLASPOAGIIPTE	717
QY	72	-----SPPISSPSYFAIPPLGSLPAELLDSPV-----LLNSSNLPSPTTGAF	114

Db 718 AVAMPIPAASPTTASPASNAVTPS---SEAKDRLQDQRQNSPAGNKENIKENETSPSF 774
Qy 115 VAOSFNWKSNGNQIIVKEEDKFSN---FSQTRSGPPASSTATYQSSNVTVQOQPW 171
Db 775 S-----KAENKGISPVVSEHRKQIDDLKFKKNDRLQPSSTSESMDQLLNKREGEKSR 828
Qy 172 SF-----QEAATKQNFSSGKGMKMTENS--SMQSPSEIASVQTNHNSNGFQSDYGNYPQ 225
Db 829 DLKDKIEPSAKDSFIENSSNCTSGSKPNSPSISLS--NTEHKGPEVTSQGVQTS 887
Qy 226 SOTLSRRSDDGYNRKYQKQVKGSE-----NPSYYKCTYPNCPKTKKVER---S 273
Db 888 SPACKQEKDDKEE-KKDAAEQVRKSTLNPNAKEFNPSFSQPKPSTTPTSPPQAQSPS 946
Qy 274 LDG--QITEI-----VYKGTIN-HPKPQNTNRNSNSSSLAIPHNSIR 314
Db 947 MVGHQOPTVYTPQVCFAPNMYVPVSPGVQPLYPIMPMPVNOAKTYRAVNNPQOR 1006
Qy 315 TEIPDOSYATH---GSGQMSAATPENSISIGDDDFEQSSQKCKSGGDEDEPDAGR 371
Db 1007 QDQHQSAMHHPASAAGPPIAATPPAYSTQVAYSPQFPNQLVQHVPHYQSQHPHVS 1066
Qy 372 WKIEGENEGMSAPGSRVREPRVVVQTTSDIDILDDGYRWRKYQKVKGNPNPRSY-- 429
Db 1067 PVIQGNARMWAPP---THAQPLVSSSATQYGAHEQTHAMVACPPLYNKETSPSFYFAI 1123
Qy 430 -----KCTHPCGVRKHVERASHDLRAVITTYEGKHNDVPAARGSG-----SHSVNRP 478
Db 1124 STGSLAQQAHPNATLHPHTP-----HPQPSATPTGQQSQHGGSH 1165
Qy 479 MPNNASHTNTAATSVLLPVIHQ-----DNSLQNRQSQAPPEGQSPF 522
Db 1166 APSVQHHQAAQALHLASPOQQAIAHYHAGLAPTPPSMTPASNTQSPNSPFAAQOTVF 1225
Qy 523 TL 524
Db 1226 TI 1227

RESULT 2
US-09-648-281-2
; Sequence 2, Application US/09648281
; Patent No. 6515197
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; TITLE OF INVENTION: Transgenic Animal Model of
; TITLE OF INVENTION: Neurodegenerative Disease and Methods of Use
; FILE REFERENCE: P-CE 4336
; CURRENT APPLICATION NUMBER: US/09/648,281
; CURRENT FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1312
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-648-281-2

Query Match 5.3%; Score 163; DB 4; Length 1312;
Best Local Similarity 19.9%; Pred. No. 9.2e-05;
Matches 120; Conservative 76; Mismatches 260; Indels 146; Gaps 22;

Qy 37 ASPLDNKPPQGGUSERTGSGVPRF--KSTPP-----PSLPL-----71
Db 658 ATPPVARTSPGGTWSVSGVRLSPKTRPRQNSIGNTPSGPVLASPOAGIIPTE 717
Qy 72 -----SPPPTSPSYFAIPGLSPAELLDSPV-----LLNSNLSPTTGAF 114
Db 718 AVAMPIPAASPTTASPASNAVTPS---SEAKDRLQDQRQNSPAGNKENIKENETSPSF 774
Qy 115 VAOSFNWKSNGNQIIVKEEDKFSN---FSQTRSGPPASSTATYQSSNVTVQOQPW 171

Db 775 S-----KAENKGISPVVSEHRKQIDDLKFKKNDRLQPSSTSESMDQLLNKREGEKSR 828
Qy 172 SF-----QEAATKQNFSSGKGMKMTENS--SMQSPSEIASVQTNHNSNGFQSDYGNYPQ 225
Db 829 DLKDKIEPSAKDSFIENSSNCTSGSKPNSPSISLS--NTEHKGPEVTSQGVQTS 887
Qy 226 SOTLSRRSDDGYNRKYQKQVKGSE-----NPSYYKCTYPNCPKTKKVER---S 273
Db 888 SPACKQEKDDKEE-KKDAAEQVRKSTLNPNAKEFNPSFSQPKPSTTPTSPPQAQSPS 946
Qy 274 LDG--QITEI-----VYKGTIN-HPKPQNTNRNSNSSSLAIPHNSIR 314
Db 947 MVGHQOPTVYTPQVCFAPNMYVPVSPGVQPLYPIMPMPVNOAKTYRAVNNPQOR 1006
Qy 315 TEIPDOSYATH---GSGQMSAATPENSISIGDDDFEQSSQKCKSGGDEDEPDAGR 371
Db 1007 QDQHQSAMHHPASAAGPPIAATPPAYSTQVAYSPQFPNQLVQHVPHYQSQHPHVS 1066
Qy 372 WKIEGENEGMSAPGSRVREPRVVVQTTSDIDILDDGYRWRKYQKVKGNPNPRSY-- 429
Db 1067 PVIQGNARMWAPP---THAQPLVSSSATQYGAHEQTHAMVACPPLYNKETSPSFYFAI 1123
Qy 430 -----KCTHPCGVRKHVERASHDLRAVITTYEGKHNDVPAARGSG-----SHSVNRP 478
Db 1124 STGSLAQQAHPNATLHPHTP-----HPQPSATPTGQQSQHGGSH 1165
Qy 479 MPNNASHTNTAATSVLLPVIHQ-----DNSLQNRQSQAPPEGQSPF 522
Db 1166 APSVQHHQAAQALHLASPOQQAIAHYHAGLAPTPPSMTPASNTQSPNSPFAAQOTVF 1225
Qy 523 TL 524
Db 1226 TI 1227

RESULT 3
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESS: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-07-741-940-7

Query Match 5.0%; Score 154; DB 1; Length 2842;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSPSLLASPLDNNKP--PQG-----GLSERTGSGVPKFKSTPPPSLPLSPPISSPS 80
Db 2206 SEISGQMKQLOANMPSISGRMTIHIPGVNRSSSTSPVSKGPKLTPASKSPSEGT 2265
QY 81 YFAIPPGLSP---AELDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
Db 2266 ATTSRGAKEPVKSEL--SFVARQTSQIGSSKAPSRGSRDSTPSRPAQQLSRPIQSP 2323
QY 120 NWKSSGCGNOQIVKEEDKFSNFSFQTRSGPPASSTATYQSSNVTVQTO-QPWSFQEA 178
Db 2324 GRNSISPRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTPSGRQMSQQLTK 2379
QY 179 QDNFSSCKGMKMTENSSMSQSPSEIASVQTHNSNGFQSDYGNVYPPQSOTLSRRSDGYN 238
Db 2380 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNGANKKVELSRMSS----- 2420
QY 239 WRKYGQKQVKGSENPRSYKCTY----PNCPTKKKVERSLDQGITVYKGTNHPKPQN 294
Db 2421 -TKSSGESDRSERPVLVROSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2472
QY 295 TRNNSNSSSLAIPHSNIRTEIPDQSYATHGSGQMS-AATPENSISIGDDDDFQSSQ 353
Db 2473 PTRSQAQTPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI----- 2514
QY 354 KCKSGGDEYDEDAKR-----WKIGENEGMSPAGSRTVREP 392
Db 2515 -----EYNDGRP-AKRHDIARSHSPRLPINRSGTWKREKSHSSLPVST----- 2562
QY 393 RVVVQTTSDIDLDGGRWKYKQVKGPNPRSYKCTHPGCPVRKHVERASHDLRAV 452
Db 2563 -----WRTGSSSILSASSESEKAKSED---EKHV-----NSI 2594
QY 453 ITTYEGKHNDVPAARG-----SGSHVNRPNPNNAHNTAATSVLLPVIHQ 502
Db 2595 SGTQSKXENQ--VSAKGTWRKIKENEPSPTNSTQTVSSGATNGAESKTLIYQWAPAVSK 2652
QY 503 S-----DNSLQNRQAPPEQSPFTLEMLQSPGSPGFGNPMQSYVNVQQLSDN 554
Db 2653 TEDVWVRIEDCPINRPSRGSPTGNTPPVIDSV-----SEKANPNIKSDKNQAKQN 2704
QY 555 V 555
Db 2705 V 2705

RESULT 4

US-08-289-548A-7
Sequence 7, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 5.0%; Score 154; DB 1; Length 2842;

Best Local Similarity 20.6%; Pred. No. 0.0017;

Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSPSLLASPLDNNKP--PQG-----GLSERTGSGVPKFKSTPPPSLPLSPPISSPS 80
Db 2206 SEISGQMKQLOANMPSISGRMTIHIPGVNRSSSTSPVSKGPKLTPASKSPSEGT 2265
QY 81 YFAIPPGLSP---AELDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
Db 2266 ATTSRGAKEPVKSEL--SFVARQTSQIGSSKAPSRGSRDSTPSRPAQQLSRPIQSP 2323
QY 120 NWKSSGCGNOQIVKEEDKFSNFSFQTRSGPPASSTATYQSSNVTVQTO-QPWSFQEA 178
Db 2324 GRNSISPRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTPSGRQMSQQLTK 2379
QY 179 QDNFSSCKGMKMTENSSMSQSPSEIASVQTHNSNGFQSDYGNVYPPQSOTLSRRSDGYN 238
Db 2380 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNGANKKVELSRMSS----- 2420
QY 239 WRKYGQKQVKGSENPRSYKCTY----PNCPTKKKVERSLDQGITVYKGTNHPKPQN 294
Db 2421 -TKSSGESDRSERPVLVROSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2472
QY 295 TRNNSNSSSLAIPHSNIRTEIPDQSYATHGSGQMS-AATPENSISIGDDDDFQSSQ 353
Db 2473 PTRSQAQTPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI----- 2514

```
QY 354 KCKSGGDEYDEBPAKR-----WKIEGENEGMSAPGSRVREP 392
Db 2515 -----EYNDGRP-AKRHDIAARSHSPRLPINRSGTWKRREKSHKSSSLPRVST----- 2562
QY 393 RVVVQTTSDIDILDDGVRWKYQKVVKGPNPRSYKCTHPGCPVRKHVERASHDLRAV 452
Db 2563 -----WRTGSSSILSASSESEKAKSED---EKHV-----NSI 2594
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRPMPNPNASHTNTAATSVRLLPVHQ 502
Db 2595 SGTQKSKENQ--VSAKGTWRKIKENEFSTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2652
QY 503 S-----DNSLQNRQSQAPPEGQSPFTLEMLQSPGFGFNGPMQSYVNVQQLSDN 554
Db 2653 TEDVWVRIEDCPINNPRSGRSPGTNTPTVIDSV-----SEKANPNIKDKDNQAKQN 2704
QY 555 V 555
Db 2705 V 2705

RESULT 5
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

```
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-452-654-7

Query Match 5.0%; Score 154; DB 1; Length 2842;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatch 235; Indels 176; Gaps 25;

QY 30 TFSFDLLASPLDNKP--PQG-----GLSRTSGVPPKPKSTPPPSPLPPPPISPS 80
Db 2206 SEISGOMKQFLOANPMSISGRMTIHPGVRNSSSTSPVSKKGPPLKTPASKSPSEG 2265
QY 81 YFAIPPLGLSP--AEILDSPVLLNSNI-----LPSPTTGAFVAQSF 119
Db 2266 ATTSRPAKPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTSRPAQPLSRPIQSP 2323
QY 120 NMKSSSGNQIQIKEEDKSFNFQTRSGPPASSTATYOSSNVTVQTO-QPMWFOEATK 178
Db 2324 GRNSISPGRNGISPPNKL-----QLPRTSSPSTASTKSSGSGKMSYTSPPGQMSQQLTK 2379
QY 179 QNFFSGKGMKMTENSSMQSPFSPETASVQTHNSGFDQSDYGNYPPOSOTLRRSDDGYN 238
Db 2380 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNGANKKVELSRMSS-- 2420
QY 239 WRKYQKQVKGSENPRSYKCTY-----PNCPTKKKVERSLDQITFVYKGTNHPKPN 294
Db 2421 -TKSGSESDRSERPVLVRQSTFIKEAPSTLRRKLEESAS-----FESLSPSRPAS 2472
QY 295 TRRNSNSSSLAIPHSNSIRTEIPDOSYATHSGGQMS-AATPENSISIGDDDFEQSSQ 353
Db 2473 PIRSQATPVLS-----PSLPMSLSTHSSVQAGWRKLPNLSPTI----- 2514
QY 354 KCKSGGDEYDEBPAKR-----WKIEGENEGMSAPGSRVREP 392
Db 2515 -----EYNDGRP-AKRHDIAARSHSPRLPINRSGTWKRREKSHKSSSLPRVST----- 2562
QY 393 RVVVQTTSDIDILDDGVRWKYQKVVKGPNPRSYKCTHPGCPVRKHVERASHDLRAV 452
Db 2563 -----WRTGSSSILSASSESEKAKSED---EKHV-----NSI 2594
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRPMPNPNASHTNTAATSVRLLPVHQ 502
Db 2595 SGTQKSKENQ--VSAKGTWRKIKENEFSTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2652
QY 503 S-----DNSLQNRQSQAPPEGQSPFTLEMLQSPGFGFNGPMQSYVNVQQLSDN 554
Db 2653 TEDVWVRIEDCPINNPRSGRSPGTNTPTVIDSV-----SEKANPNIKDKDNQAKQN 2704
QY 555 V 555
Db 2705 V 2705

RESULT 6
US-08-449-731-7
; Sequence 7, Application US/08449731
; Patent No. 6413727
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
```


QY 81 YFAIPPLGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
Db 2267 ATTSPPRAKPSVKSEL--SPVARTSOIGSSKAPSGSRDSTPSRPAQQLSRPIQSP 2324
QY 120 NWKSSGNGNOQIVKEEDKSFNFSTQTRSGPPASSTATYQSSNVTVQTO-QPWSFQEAATK 178
Db 2325 GRNSISPRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTSPPGRQMSQQNLTK 2380
QY 179 QDNFSSGCKGMKWTENSSMSQSFPEIASVQTNHNSGFQSDYGNYPQSQTLRSRSDGYN 238
Db 2381 QTGLS-----KNASSIP--RSESASGKLQNMNN-----GNGANKVELSRMSS----- 2421
QY 239 WRKYGQKVGKSENPRSYKCTY-----PNCPTKKKVERSLDGOITTEIVYKGTNHPKPQN 294
Db 2422 -TKSSGSESDRSERPVLVRQSTFIKEAPSTLRRKLEESAS-----PESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSIRTEIPDQSYATHSGQWDS-AATPENSISIGDDDDFEQSSQ 353
Db 2474 PTRSQATPVLS-----PSLPDMSLTHSSVQAGGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEPDAGR-----WKIEGENEGMSAPGSRVTREP 392
Db 2516 -----EYNDGRP-AKRHDIAKSHSESPRLPINRSGTWKREHSHSSSLPRVST----- 2563
QY 393 RVVVQTTSDIDLDGVRWRKYGQKVVKGPNPRSYKCTHPGCPVKRKHVERASHDLRAV 452
Db 2564 -----WRTGSSSILSASSESEKAKSED-----EKHV-----NSI 2595
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRPMNPNNASHTNTAATSVRLLPVHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENBFPSTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLONQSOAPPEQSPFTLEMLQSPGFGFGNPMQSVNQQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSDKNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 8

US-08-289-548A-2
; Sequence 2, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A

; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9299
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-289-548A-2

Query Match 5.0%; Score 154; DB 1; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSFSDLLASPLDNKBP--PQG-----GLSERTSGVPKPKSTPPSLPLSPSPSPS 80
Db 2207 SEISGQMKQFLOANMPSISGRRTMIHIPGVRNSSSTSPVSKGPKPLKTPASKSPSEQT 2266
QY 81 YFAIPPLGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
Db 2267 ATTSPPRAKPSVKSEL--SPVARTSOIGSSKAPSGSRDSTPSRPAQQLSRPIQSP 2324
QY 120 NWKSSGNGNOQIVKEEDKSFNFSTQTRSGPPASSTATYQSSNVTVQTO-QPWSFQEAATK 178
Db 2325 GRNSISPRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTSPPGRQMSQQNLTK 2380
QY 179 QDNFSSGCKGMKWTENSSMSQSFPEIASVQTNHNSGFQSDYGNYPQSQTLRSRSDGYN 238
Db 2381 QTGLS-----KNASSIP--RSESASGKLQNMNN-----GNGANKVELSRMSS----- 2421
QY 239 WRKYGQKVGKSENPRSYKCTY-----PNCPTKKKVERSLDGOITTEIVYKGTNHPKPQN 294
Db 2422 -TKSSGSESDRSERPVLVRQSTFIKEAPSTLRRKLEESAS-----PESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSIRTEIPDQSYATHSGQWDS-AATPENSISIGDDDDFEQSSQ 353
Db 2474 PTRSQATPVLS-----PSLPDMSLTHSSVQAGGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEPDAGR-----WKIEGENEGMSAPGSRVTREP 392
Db 2516 -----EYNDGRP-AKRHDIAKSHSESPRLPINRSGTWKREHSHSSSLPRVST----- 2563
QY 393 RVVVQTTSDIDLDGVRWRKYGQKVVKGPNPRSYKCTHPGCPVKRKHVERASHDLRAV 452
Db 2564 -----WRTGSSSILSASSESEKAKSED-----EKHV-----NSI 2595
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRPMNPNNASHTNTAATSVRLLPVHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENBFPSTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLONQSOAPPEQSPFTLEMLQSPGFGFGNPMQSVNQQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSDKNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 9

US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-654-2

Query Match 5.0%; Score 154; DB 1; Length 2843;
Best Local Similarity 20.8%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TFSFDLLAGPLDNNKP--PQG-----GLSERTGSGVPKFKSTPPPLSLPPSPISPS 80
DB 2207 SEISGQMKQPLQANMPSISGRWTMIHPGVNRSSSTSPVSKGPKLPKTPASKSPSEGT 2266
QY 81 YFAIPGCLGP---AELDGSVLNLSNI-----LPSFTTGAFVAQSF 119
DB 2267 ATTSPRGAKPSVKSEL--SPVARQTSQIGSSKAPSRSGRSDTSPRPAQQLSRDIQSP 2324
QY 120 NWKSSGGNQIIVKEDKSFNSFQTRSGPPASSTATVQSSNVVTQTO-QPWSFQEA TK 178
DB 2325 GRNSISPGRNGISPPNKLKLS---QLPRTSPSTASTKSGSGKMSVTSPGRQMSQQLTK 2380
QY 179 QDNFSSGKGMKWTENSSSQSPFEIASVQTNHNGFQSDYGNYPQSQOTLSRRSDGYN 238
DB 2381 QTGLS-----KNASTIP--RSEASKGLNOMN-----NGANKVELSRMSS----- 2421
QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKYVERSLDQGITBIVYKGTNNHFKPQN 294
DB 2422 -TKSSGESDRSRPVLVRQSTTIKEAPSLTRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSINTETPDQSYATHGSGQWDS-AATPENSISIGDDDFQSSQ 353
DB 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGWRKLPNLSPTI----- 2515

RESULT 10

US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

QY 354 KCKSGGDEYDEPDPAKR-----WKTEGENEGMSAPGSRVTRREP 392
DB 2516 -----EYNDGRP-AKRHDIAHSHSESPSRLPINRSGTWKREHSHKSSSLPRVST----- 2563
QY 393 RVVVQTTSDIDLDDGYRKYGQKVVKGNPNRPSYKCTHFGCPCVRKHVERASHDLRAV 452
DB 2564 -----WRRTGSSSSILSASESESEKAKSED-----EKHV-----NSI 2595
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRMPNPNASNTHTNTATSVRLLPVIHQ 502
DB 2596 SCTKQSKENQ--VSAGTWRKIKENEFSPNTSQTSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLQNRQOAPPEGQSPFTLEMLQSPGSPGFGNPMQSYVNOQLSDN 554
DB 2654 TEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSV-----SEKANPNIKDSKDNQAKQN 2705
QY 555 V 555
DB 2706 V 2706

MOLECULE TYPE: protein

US-08-452-655B-2

Query Match 5.0%; Score 154; DB 1; Length 2843;

Best Local Similarity 20.6%; Pred. No. 0.0017; Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

```

QY 30 TFSDDLASPLDNKP--PG-----GLSERTGCVKFKSTPPPSLPLSPSPSPS 80
DB 2207 SEISGQKQLOANMPSISGRRTMIHIPGVNRSSSTSPVSKGPKLTPASKSPSEGQT 2266
QY 81 YFAIPPLGSP---AELDSPLVLLNSNI-----LPSPTTCAFAVQSF 119
DB 2267 ATTSRGAKEPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTPSRPAQQLSRFIQSP 2324
QY 120 NWKSSGNGNQIIVKEEDKSFNFQTRSGPPASSTATYQSSNVTVQTO-QPMSFQEAATK 178
DB 2325 GRNSISPGRNGISPPNKL-----QLPRTSSPSTASTKSSGSGKWSYTSFGQMSQQLTK 2380
QY 179 QDNFSSGKGMKKTENSSMQSPFETASVQTNHNGFQSDYGNYPQSQTLRRSDDGYN 238
DB 2381 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNANKKVELSRMSS----- 2421
QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKKVERSILDGOITEIVYKGTNHPKPQN 294
DB 2422 -TKSSGESDRSRPVLVRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSIRTEIPDOSYATHGSGQMDs-AATPENSISIGDDDDFEQSSQ 353
DB 2474 PTRSQAQTPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEPDADR-----WKIEGENEGMAGPSRTVREP 392
DB 2516 -----EYNDGRP-AKRHDIAHSHSESPRLPINRSGTWKREHSHSSLPVST----- 2563
QY 393 RVVVQTTSDIDILDDGVRWKYQKVGKGNPNRSPRYKCTHPCGCPVKRKHVERASHDLRAV 452
DB 2564 -----WRTGSSSILSASSESEKAKSED---EKHV-----NSI 2595
QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNNAHNTNTAATSVLLPVIHQ 502
DB 2596 SGTQKQKENGQ--VSAKGTWRKIKENEPSTNSTSQTVSSGATGAESKTLIYQMAPAVSK 2653
QY 503 S-----NSLQNRQSQAPPEQSPFTLEMLQSPGSGFGNGNPMQSYVQOQLSDN 554
DB 2654 TEDVWVRIEDCPINPRSGRSPTNTPPVIDSV-----SEKANPNIKDKDNQAKQN 2705
QY 555 V 555
DB 2706 V 2706

```

RESULT 11

US-08-452-655B-7
Sequence 7, Application US/08452655B

Patent No. 5783666

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERTS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 1001 G Street, NW

```

; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: YES
; ANTI-SENSE: NO
; US-08-452-655B-7

```

Query Match 5.0%; Score 154; DB 1; Length 2843;

Best Local Similarity 20.6%; Pred. No. 0.0017;

Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

```

QY 30 TFSDDLASPLDNKP--PG-----GLSERTGCVKFKSTPPPSLPLSPSPSPS 80
DB 2207 SEISGQKQLOANMPSISGRRTMIHIPGVNRSSSTSPVSKGPKLTPASKSPSEGQT 2266
QY 81 YFAIPPLGSP---AELDSPLVLLNSNI-----LPSPTTCAFAVQSF 119
DB 2267 ATTSRGAKEPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTPSRPAQQLSRFIQSP 2324
QY 120 NWKSSGNGNQIIVKEEDKSFNFQTRSGPPASSTATYQSSNVTVQTO-QPMSFQEAATK 178
DB 2325 GRNSISPGRNGISPPNKL-----QLPRTSSPSTASTKSSGSGKWSYTSFGQMSQQLTK 2380
QY 179 QDNFSSGKGMKKTENSSMQSPFETASVQTNHNGFQSDYGNYPQSQTLRRSDDGYN 238
DB 2381 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNANKKVELSRMSS----- 2421
QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKKVERSILDGOITEIVYKGTNHPKPQN 294
DB 2422 -TKSSGESDRSRPVLVRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSIRTEIPDOSYATHGSGQMDs-AATPENSISIGDDDDFEQSSQ 353
DB 2474 PTRSQAQTPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEPDADR-----WKIEGENEGMAGPSRTVREP 392
DB 2516 -----EYNDGRP-AKRHDIAHSHSESPRLPINRSGTWKREHSHSSLPVST----- 2563
QY 393 RVVVQTTSDIDILDDGVRWKYQKVGKGNPNRSPRYKCTHPCGCPVKRKHVERASHDLRAV 452
DB 2564 -----WRTGSSSILSASSESEKAKSED---EKHV-----NSI 2595

```

QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNPNASHTNTAATSVRLLPVHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENEFPTNSTQTSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLONORSQAPPEQSPPTLEMLQSPGSGFGNPMQSVYVNOOQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSDKDNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 12

US-08-370-235A-2
; Sequence 2, Application US/08370235A
; Patent No. 5910418
; GENERAL INFORMATION:
; APPLICANT: VOGELSTEIN, BERT
; APPLICANT: KINZLER, KENNETH W.
; APPLICANT: HILL, DAVID E.
; APPLICANT: JOHNSON, KAREN A.
; TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DERMING
; TITLE OF INVENTION: MUTATIONS IN THE APC GENE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & WITCOFF, LTD.
; STREET: 1001 G STREET, N.W.
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20001

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,235A
; FILING DATE: 01-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KAGAN, SARAH A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.48688
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 508 9100
; TELEFAX: 202 508 9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-370-235A-2

Query Match 5.0%; Score 154; DB 2; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;
QY 30 TSFSDLLASPLDNNKP--PQG-----GLSERTGCVKFKSTPPSLPLSPPIPSPS 80
Db 2207 SEISGQMKQPLQANMPSISRGRTMIHIPGVNRSSSTSPVSKGPPPLKTPASKSPSGQT 2266
QY 81 YFAITPGGLSP---AELLDSPLVLLNSNI-----LPSPTTGAFVAQSF 119
Db 2267 ATTSPRGAKPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTPSRPAQQPLSRPIQSP 2324
QY 120 NWKSSSGNQIVKEEDKSFNSFQTRSGPPASSTATYOSSNVTVQTO-QPKWSFQEA TK 178
Db 2325 GRNSISPRNGISPPNKLK-----QLPRTSPFSTASTKSSGSKGWSYTSPPGRQWSQQLTK 2380
QY 179 QDNFSSGKGMWKTENSSSMQSFPEIASVQTNHNSGFGQSDYGNYPQOSQTLRRSRDDGYN 238

Db 2381 QTGLS-----KNASSIP--RSESASKLQNMN-----GNANKKVELSRMSS-- 2421
QY 239 WRKYGQKQVKGSENPRSYKCTV---PNCPTKKKVERSILDGOITEIVYKGTNHPKPN 294
Db 2422 -TKSSGSESDRSPVLRQSTFIKEAPSPTRLRKLEBAS-----FESLSPSRPAS 2473
QY 295 TRSNSNSSSLAIPHSNIRTEIPDOSYATHSGQMDS-AATPENSISIGDDDDFEQSSQ 353
Db 2474 PTRSQATPVL-----PSLPDMSLSLTHSSVQAGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEPDADR-----WKIEGENEGHSAFGSRTVREP 392
Db 2516 -----EYNDGRP-AKRHDIAHSHSESPRLPINRSGTWKREHSHSSSLPRVST-- 2563
QY 393 RVVVQTSDIDLDDGVRWKYQKVVKGPNPRSYKCTHPCGCPVRKHVERASHDLRAV 452
Db 2564 -----WRTGSSSSILSASESESEKAKSED---EKHV-----NSI 2595
QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNPNASHTNTAATSVRLLPVHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENEFPTNSTQTSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLONORSQAPPEQSPPTLEMLQSPGSGFGNPMQSVYVNOOQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSDKDNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 13

US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940

; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A. 32,141
; REGISTRATION NUMBER: 1107.49964
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-582-2

Query Match 5.0%; Score 154; DB 3; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSFSDLLASPLDNNKP--PG-----GLSERTGSGVPKFKSTPPPSLPLSPSPSS 80
DB 2207 SEISGQMKQPLQANMPSISGRTHIHPGVNRSSSTSPVSKGPKLTPASKSPSEGQT 2266

QY 81 YFAIPPGGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
DB 2267 ATTSFRAKPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTPSRPAQPLSRPIQSP 2324

QY 120 NWKSSSGGNQOIVKEEDKFSNFSQTRSGPPASSTATYQSSNVTVQTQ-OPMSFQBATK 178
DB 2325 GRNSISPGRNGISPPNKLKLS---QLPRTSPSTASTKSGSGKMSYTPSGRQMSQNLTK 2380

QY 179 QDNFSSGKGMKMTENSSSMQSPFPIASVQTNHNGFQSDYGNYPQSQTLRSRDDGYN 238
DB 2381 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNANKKVELSRMSS----- 2421

QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKKVERSLODQITETIVYKGTNNHPKPN 294

QY 2422 -TKSGSESDSERPVLVRSQSTFIKEAPSLRKLSEAS-----FESLSFSPSRPAS 2473

QY 295 TRNNSNSSLAIPHNSIIRTEIPQSYATHGSGQMS-AATPNSISISIGDDDFEQQS 353
DB 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGWRKLPNLSPTI----- 2515

QY 354 KCKSGGDEYDEPDAR-----WKIEGENGMAPGSRVREP 392
DB 2516 -----EYNDGRP-AKRHDIAHSHSPSRPLPINRSGTWKREHSKSSSLPRVST----- 2563

QY 393 RVVVQTTSDIDLDGYRWKYGQVKGPNPRSYKCTHPGCPVRKHVERASHDLRAV 452
DB 2564 -----WRRTGSSSILSASSESEKAKSED---EKHV-----NSI 2595

QY 453 ITTYEGKHNDVPAARG-----SGHSVNRMPNPNASNTNTAATSVRLLPVIHQ 502
DB 2596 SGTQSKXENQ--VSAKGTWRKI KENEFSPNTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2653

QY 503 S-----DMSLQNRQSQAPPEQSPFTLEMLQSPGSGFGSGNPMQSYVNOQQSLDN 554
DB 2654 TEDVMVRIEDCPINNPRSGRSPTGNTTPVIDSV-----SEKANPNIKDSKDNQAKQN 2705

QY 555 V 555
DB 2706 V 2706

RESULT 14

US-08-450-582-7
; Sequence 7, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY

Query Match 5.0%; Score 154; DB 3; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSFSDLLASPLDNNKP--PG-----GLSERTGSGVPKFKSTPPPSLPLSPSPSS 80
DB 2207 SEISGQMKQPLQANMPSISGRTHIHPGVNRSSSTSPVSKGPKLTPASKSPSEGQT 2266

QY 81 YFAIPPGGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
DB 2267 ATTSFRAKPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTPSRPAQPLSRPIQSP 2324

QY 120 NWKSSSGGNQOIVKEEDKFSNFSQTRSGPPASSTATYQSSNVTVQTQ-OPMSFQBATK 178
DB 2325 GRNSISPGRNGISPPNKLKLS---QLPRTSPSTASTKSGSGKMSYTPSGRQMSQNLTK 2380

QY 179 QDNFSSGKGMKMTENSSSMQSPFPIASVQTNHNGFQSDYGNYPQSQTLRSRDDGYN 238
DB 2381 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNANKKVELSRMSS----- 2421

QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKKVERSLODQITETIVYKGTNNHPKPN 294

Db 2422 -TKSGSESDRSPVLRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRNSNSSSLAIPHSNIRTEIPDOSYATHGSGQMDs-AATPENSISIGDDDDFQSSQ 353
Db 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI-----2515
QY 354 KCKSGGDEYDEBDAKR-----WKIEGENEGMSAPGSRVREP 392
Db 2516 -----EYNDGRP-AKRHDIARSHSESPSRLPINRSGTWKREHSHSSSLPRVST----2563
QY 393 RVVVQTTSDIDLDGVRWKYQKVKVGNPNRPSYKTHPCGCPVRKHVERASHDLRAV 452
Db 2564 -----WRTGSSSSILSASSESEKAKSED---EKHV-----NSI 2595
QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNNASNHTTAATSVRLLPVIHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENEFPTNSTQTVSSGATNGAESKTLIYQMAFAVK 2653
QY 503 S-----DNSLQNRQAPPEGQSPFTLEMLQSPGSGFGFGNPMQSYVNOQQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSKDNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 15

US-08-449-731-2
; Sequence 2, Application US/08449731
; Patent No. 6413727
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; ANAND, RAKESH
; CARLSON, MARY
; GRODEN, JOANNA
; HEDGE, PHILIP J.
; JOSLYN, GEOFF
; KINZLER, KENNETH
; MARKHAM, ALEXANDER F.
; NAKAMURA, YUSUKE
; THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,731
; FILING DATE: 25-May-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,548
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-449-731-2
Query Match 5.0%; Score 154; DB 4; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;
QY 30 TFSDLLASPLNNKP--PQG-----GLSRTSGVGVKFKSTPPPSPLSPSPSS 80
Db 2207 SBITSGMKQPLQANMPSISRGTMIHPCVRNSSSTSPVSKGPKLTPASKSPSEGOT 2266
QY 81 YFAIPPLGSP---AELLDSPVLLNSNI-----LPSPTTCFAVQSF 119
Db 2267 ATTSPRGAKPSVKSEL--SPVARQTSQIGSSKAPSRSGRSDTSPRPAQQLSRFIQSP 2324
QY 120 NMKSSGNGQIQIVKEEDKSFNSFOTRSGPPASSTATYQSSNVTVQTQ--QPMWFOEATK 178
Db 2325 GRNISTPGRNGISPPNKLs-----QLPRTSSPSTASTKSSGSGKMSYTSFGRQMSQQLTK 2380
QY 179 QDNFSGKGMKMTENSSMQSPFETASVQTNHNGFQSDYGNYPQSQTLRRRSDGYN 238
Db 2381 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNANKKVELSRMSS---2421
QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKKVERSLDGOITEIVYKGTNHPKQN 294
Db 2422 -TKSGSESDRSPVLRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRNSNSSSLAIPHSNIRTEIPDOSYATHGSGQMDs-AATPENSISIGDDDDFQSSQ 353
Db 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI-----2515
QY 354 KCKSGGDEYDEBDAKR-----WKIEGENEGMSAPGSRVREP 392
Db 2516 -----EYNDGRP-AKRHDIARSHSESPSRLPINRSGTWKREHSHSSSLPRVST----2563
QY 393 RVVVQTTSDIDLDGVRWKYQKVKVGNPNRPSYKTHPCGCPVRKHVERASHDLRAV 452
Db 2564 -----WRTGSSSSILSASSESEKAKSED---EKHV-----NSI 2595
QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNNASNHTTAATSVRLLPVIHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENEFPTNSTQTVSSGATNGAESKTLIYQMAFAVK 2653
QY 503 S-----DNSLQNRQAPPEGQSPFTLEMLQSPGSGFGFGNPMQSYVNOQQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSKDNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

Search completed: January 20, 2004, 16:18:25
Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:19:31 ; Search time 38 Seconds
(without alignments)
3094.123 Million cell updates/sec

Title: US-09-890-811B-10
Perfect score: 3050
Sequence: 1 MASSSGSLDTSASANSFTNF.....FSSRTKEPRDDMFLESILIC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1605	52.6	560	9	US-09-810-264-10
2	1145	37.5	519	16	US-10-278-536-182
3	906	29.7	393	15	US-10-278-173-88
4	899	29.5	557	11	US-09-934-455-326
5	899	29.5	557	12	US-10-302-267-198
6	888.5	29.1	278	9	US-09-810-264-24
7	884	29.0	521	15	US-10-278-173-36
8	884	29.0	521	16	US-10-278-536-34
9	861.5	28.2	514	12	US-10-225-068-140
10	861.5	28.2	514	15	US-10-295-403-82
11	858	28.1	577	9	US-09-810-264-24
12	852	27.9	687	11	US-09-934-455-18
13	852	27.9	687	15	US-10-295-403-76
14	848.5	27.8	513	15	US-10-295-403-78
15	753	24.7	266	12	US-10-259-165-64

16	742	24.3	568	11	US-09-934-455-10
17	742	24.3	568	12	US-10-225-068-70
18	643.5	21.1	623	12	US-10-259-165-196
19	625.5	20.5	406	9	US-09-810-264-12
20	619.5	20.3	413	9	US-09-810-264-14
21	618.5	20.3	583	9	US-09-810-264-38
22	606	19.9	430	9	US-09-810-264-16
23	460.5	15.1	258	9	US-09-810-264-32
24	353	11.6	173	9	US-09-810-264-22
25	318	10.4	191	11	US-09-934-455-54
26	316.5	10.4	528	11	US-09-934-455-406
27	315.5	10.3	318	11	US-09-934-455-362
28	315.5	10.3	318	15	US-10-278-173-42
29	310.5	10.2	553	11	US-09-533-029-90
30	310.5	10.2	553	11	US-09-934-455-360
31	309.5	10.1	326	11	US-09-934-455-62
32	309.5	10.1	536	11	US-09-934-455-58
33	309.5	10.1	536	15	US-10-278-173-40
34	309.5	10.1	536	16	US-10-278-536-78
35	305.5	10.0	528	12	US-10-225-068-126
36	300	9.8	337	11	US-09-934-455-296
37	300	9.8	337	15	US-10-278-173-90
38	295	9.7	102	9	US-09-810-264-18
39	288	9.4	277	12	US-10-225-068-144
40	280.5	9.2	147	11	US-09-533-029-16
41	280.5	9.2	147	15	US-10-278-173-64
42	272.5	8.9	317	11	US-09-934-455-12
43	268	8.8	195	11	US-09-934-455-460
44	264.5	8.7	489	11	US-09-533-029-60
45	259.5	8.5	179	11	US-09-934-455-56

ALIGNMENTS

RESULT 1

US-09-810-264-10
; Sequence 10, Application US/09810264
; Patent No. US200207675A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Virginia C.
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; APPLICANT: Zhang, Lingyu
; TITLE OF INVENTION: WRKY Transcription Factors and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: 1183
; CURRENT APPLICATION NUMBER: US/09/810,264
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,467
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Helianthus annuus
; US-09-810-264-10

Query Match 52.6%; Score 1605; DB 9; Length 560;
Best Local Similarity 57.0%; Pred. No. 1.4e+106;
Matches 335; Conservative 77; Mismatches 120; Indels 56; Gaps 16;
Qy 13 SANSFTNFTFTSTHPTTTSFSDLLASPLDNKKPQ-----GGLSERTGSGVPKFKSTPPP 67
Db 2 SFSSSSGITLETTPSPSPSFMSFSFSDQPPPTTGLAARIAERVGSGIPKFKSTPPP 61
Qy 68 SLPLSPPTSPSSVFAIPGLSPAEILLDSPLVINSNLLPSPPTTGAFVAGSQSNWKKSSG- 126
Db 62 SLPIPSPAVSSPSYFAIPAGLSPAEILLDSPLVINSNLLPSPPTTGAFVAGSQSNWKKSSG- 121
Qy 127 --GNQIVKEDKDSFNSFQTRSGPPASSTATYQSSNVTVTQQPMSFOEATKQDNFSS 184

Db 122 FHNEHSIKKEQSLADFSRPLQHHPE-----QIWNQ-----KQIQDQD 164
Qy 185 GKGMKTKENSSSQSPSPETASVQTHNSNGFQSDYGNYPPOS-----QTLRRSDGYN 238
Db 165 EKSLTQSGHSPPMQSPSPETATQDSNQASQFQSDYTNSSSNFNQTLQKKSDEGYN 224
Qy 239 WRKYGQKVGKSENPRSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHPQONTNRN 298
Db 225 WRKYGQKVGKSENPRSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHPQONTNRN 294
Qy 299 SSNSSS-----LAIPHNSIRTEIPDOSYATHGSGQMSDAAATPENSISISIGDDFEQSSQK 354
Db 285 SSSASNTLQMSQASSNHDVDPQSYVSHGSGQVDSVTPPENSISISVGDDEFDRS--- 341
Qy 355 CKSGGD--EYDEDEPAKRWKIEGENEGMS-APGSRVREPRVVQTTSDIDILDDGYRW 411
Db 342 -RSGDGVTVDEDEPAKRWKV-SENEGISMIGTKTVREPRVVQTTSDIDILDDGYRW 399
Qy 412 RYKGVKVGKGNPNRPSYKCTHPCPVKVERASHDLRAVITYGKKNHDPVPAARGSG 471
Db 400 RYKGVKVGKGNPNRPSYKCTHPCPVKVERASHDLRAVITYGKKNHDPVPAARGSG 459
Qy 472 SHSVNRPMPNNSNHTNTAATSVRLPLVH-QSDNSL--QNORSQAPP--EGQSPFTLEM 526
Db 460 HRLQASTLSNNAWMT-----TKPMALSHYQVDNSVDPTRGPRYPSPSENQAPFTLEM 513
Qy 527 LQSPGSGFGGFGNPMQSYVNOQOLSDNVFSRRTKEEPRDDMFLESLL 574
Db 514 LQSSDNPKYRFENALKSNYNEHN-SRRTFST-TKEEPRDDMFLESLL 559

RESULT 2

US-10-278-536-182
; Sequence 182, Application US/10278536
; Publication No. US20030131386A1
; GENERAL INFORMATION:
; APPLICANT: Samaha, Raymond
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Keddie, James
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MEI-011
; CURRENT APPLICATION NUMBER: US/10/278,536
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 182
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G915
US-10-278-536-182

Query Match 37.5%; Score 1145; DB 16; Length 519;
Best Local Similarity 43.68; Pred. No. 1.1e-73;
Matches 260; Conservative 82; Mismatches 154; Indels 100; Gaps 15;
Qy 1 MASSGSLD---TSASANSFTNFTTFFPMTTFSDDLASPLDNNKPPQGLSERTGSG 57
Db 1 MAASFLTMDSRTRQNNMGNSANMQSGRTSTSSLEDL-----E 39

Qy 58 VPKFKSTPPPSLPLSPPIPSISSYFAIPPLGSLPAEILLDSVLLNSS-NILPSPPTTGAFVA 116
Db 40 IKFRSFASSISISPSLVSPTCP-----SPSLFLDSPAFVSSSANVLASPTTGALI- 92
Qy 117 QSNFKWSSGGNQIIVKEEDKFSN-----FSQTRSGPPASSATATYOSSNVTVTQGP 170
Db 93 -----TWVNTQKGINEGDKSNNNFNLFDFSPHTQSSGVAPTTTTTTTTTTTNS- 144
Qy 171 WSPQBATKQDNFSSGKMMKTKENSSSQSPSPETASVQTHNSNGFQSDYGNYPPOSQTL 230
Db 145 -SIFOSQEQKKNQSEQMSQETRPNNAVS-----YNGRE-----Q 180
Qy 231 RRSDDGYNRWKYKQKVGKSENPRSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHP 290
Db 181 RXGEDGYNRWKYKQKVGKSENPRSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHP 240
Qy 291 KPQNTNRNSNSSSL-----AIPHNSIRTEIPDOSYATHGSGQMSDAAATPENSISI 343
Db 241 KPQSTRSSSSSTFSAVYNASLDHNRQASSDQPNNSNNSFHQSDSFGMQQEDNTTSDSV 300
Qy 344 GDDDFEQQSKCKSGDEYDEDEPAKRWKIEGENEGMSAPGSRVREPRVVQTTSDID 403
Db 301 GDDFEQQS-SIVSRDEEDCGSEPEAKRWKGNETNGNGGSGKTVREPRVVQTTSDID 359
Qy 404 ILDDGYNRWKYKQKVGKGNPNRPSYKCTHPCPVKVERASHDLRAVITYGKKNHD 463
Db 360 ILDDGYNRWKYKQKVGKGNPNRPSYKCTHPCPVKVERASHDLRAVITYGKKNHD 419
Qy 464 VPAARGSGSHSVNRPMPNNSNHTNTAATSVRLPLVHOSDNLQNRQAPGEGSPFT 523
Db 420 VPAARGSG-YATNRAPOQSSSVPIRPAATA-----GHSNYTTSSQAPYT 462
Qy 524 LEMIQ-----SPGSGFGGFGNPMQSYVNOQOLSDNVFSRRTKEEPRDD-MELESLL 574
Db 463 LQMLHNNTNTGTFGYAMNNNNNNNNLQTOQNFVGGFSGRAKEEPEETSFDFSPM 518

RESULT 3

US-10-278-173-88
; Sequence 88, Application US/10278173
; Publication No. US20030061637A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Pineda, Omaira
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Keddie, James
; APPLICANT: Heard, Jacqueline
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
; FILE REFERENCE: MEI-009
; CURRENT APPLICATION NUMBER: US/10/278,173
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US/09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 88
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G177
US-10-278-173-88


```
; NAME/KEY: DOMAIN
; LOCATION: (227)....(285)
; OTHER INFORMATION: Conserved domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (407)....(465)
; OTHER INFORMATION: Conserved domain
US-10-225-068-140

Query Match      28.2%; Score 861.5; DB 12; Length 514;
Best Local Similarity 38.3%; Pred. No. 2.2e-53;
Matches 217; Conservative 62; Mismatches 139; Indels 149; Gaps 21;

QY 31 SPSDLASPLDNKPPQGLS-----ERTGSV-----PKFKSTPPSLPLSPPP 75
DB 60 SFSQLLAGAGSSPATAAAAAATASYQRLGETNSSSGDVPFRKQNRPTGLMISQSQ 119
QY 76 ISPSYFAIPPGSLPAELDSPVLLNSSNLPSTTGAFVAQSFNWKSSSGNQIIVKEE 135
DB 120 -SP-SMFTVPGLSPAMLLDSPFLG---LFSVQGSY-----GMTHQQAQA- 162
QY 136 DKSFNFSFOTRSGPPASATYQSSNVTVQOPMSFQEAATKQDNFSSGKGMKNTENS 195
DB 163 -----VTAQAOANANMQPQTEYP-----PPSQVQSFSSGQAQIPTSA 202
QY 196 SMQSPSPFIASVQTNHNSGFQSDYGNYPQSQTLRSBDDGYNWKYQKQVKGSENP 255
DB 203 PAQRETSDVITIE-----HRSQPLNVDPKADDDGYNWKYQKQVKGSEFP 250
QY 256 YKCTYPCNCTKKKVERSLEDQITEIVYKGTNNHPKQNTNR--NSSNSSSLAIPHNSIR 314
DB 251 YKCTNPGCPVKKVERSLEDQVTEIYKQHNHPPQNTKRGKNDKNTANI---NGSSIN 307
QY 315 TEIPDOSYATHGQMSDAATPENSISIGDDDFEQSSQK-----KSGDEY----- 362
DB 308 N-----NRGSELGASQFQTNSSNKTREQHEAVSQATTTTEHLSEASDGEEVNGET 359
QY 363 -----DEDEPDAKRWKIEGE-NEGMSAPGSRVREPRVVVQTTSDIDLDGGRWKYQ 416
DB 360 DVREKDENEPDKRRSTREVRISAPAAASHRTVTEPRIIVQTTSEVDLLDDGGRWKYQ 419
QY 417 KVKGNPNRPSYKCTHPCGPKVRKHVERASHDLRAVITTYEGKHNHDPVAARGSGSHSVN 476
DB 420 KVKGNPNRPSYKCTHPCGPKVRKHVERAATDPKAVVTTVEGKHNHDLPAK--SSH--- 475
QY 477 RMPNNASNHTNTAATSVRLLPVTHQSDNSLQNSQAPPEGOSPTFLEMLQSPGSGF 536
DB 476 -----AAAAQLRP-----DN-----RPG----- 489
QY 537 GFQGNPMQSYVNOOQLSDNVFSSRTKEE 563
DB 490 GLAN-----LNQOQOQOQPVARLRKEE 511

RESULT 11
US-09-810-264-24
; Sequence 24, Application US/09810264
; Patent No. US20020076775A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Virginia C.
```

```
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; APPLICANT: Zhang, Lingyu
; TITLE OF INVENTION: WRKY Transcription Factors and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: 1183
; CURRENT APPLICATION NUMBER: US/09/810,264
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,467
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Glycine max
US-09-810-264--24

Query Match      28.1%; Score 858; DB 9; Length 577;
Best Local Similarity 35.8%; Pred. No. 4.5e-53;
Matches 216; Conservative 78; Mismatches 175; Indels 134; Gaps 20;

; 41 DNKKPOGGLSERTGSGVPKFKSTPPPSLPSPSPSSYFAIPGCLSPAEILLDSFVLL 100
; 10 DPNRNSAADAAPAGA-RYKLLSPAKLPISRSPC-----VTISPCLSPTSFLESPVLL 63
; 101 NSSNIIPLSPPTG--AFVAQSFNWKSSSGGNQOI-----VKEEDKSFNFSFQTR 147
; 64 SNMKVPSPTGSLSLHHTAYGSMTSAASATFPVLCASIAIPLMRENIAF--FEFKPH 121
; 148 SGPPASSTATYQSSNVTVTQQPWMSFOEATKQNFSGSK-----GMKMTENSSMQ-- 198
; 122 SG-----SNMVPADFNDHASEKSTQID--SQKAAQAFDSSALVKNESASPSNEL 168
; 199 SFSPETASVQTNHNSGQSDYGNYPQSQ-----TLRSRSDGQYXW 239
; 169 SLSPVKWIAQEAASARVEGDLDELNPRSNITTTGLQASQVDNRGSLTVAARVSDDGYNW 228
; 240 RYKGQKQVKGSENPSPRYKCTYPCNPTKKKVERSLDGOITEIVYKGTNNHPKQNTNRS 299
; 229 RYKGQKHVGSEFPSPRYKCTHPNCEVKKLFERSHDGQITEIYKGTHTDHPQPNRRYS 288
; 300 SNSSSLAIPHNSIRTEIPDQ-----SYATHGSGQMDSAATPENSISIGDDDFEQQSQK 354
; 289 A-CTIMSVOEDRDKASLTSRDDKGSNMCCGGHLEAPDCKPELLPVATNDGDLGLVL 347
; 355 CKSGGDEYDEPDAPKRWKIEGENGMS--APGSRVTREPRVVVQTTSDIDLDGQYRW 412
; 348 SNRNDEVDVDDPFSEKSKMD--VGIADITPVVKPIREPRVVVQTLSEVDILDDGQYRW 404
; 413 KYGQKVKGNPNRSPRYKCTHPGCPVRKHVERASHDLRAVITYYEGKNNHDPVAAARGSGS 472
; 405 KYGQKVRGNPNRSPRYKCTNTGCPVRKHVERASHDPAVITYYEGKNNHDPVTAARN-C 463
; 473 HSNVRNPMNNASNHT-----NTAATSVRL-----PVTHQS 503
; 464 HDWAGFA--SASGQTRVPRPESDTISLDLGMGISPAAENTNSQGRNMLSEFGDSQIHTS 521
; 504 DNSLQONORSQAPPEGQSPFTLEMLQSPGFGFGNGPMOSYVYVQOQSDNVSFSTRKEE 563
; 522 NSNFK-----FVHTTAPGYFGV-----LNNNSNPYGSKEN 552
; 564 PRD 566
; 553 PSD 555

RESULT 12
US-09-934-455-18
; Sequence 18, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omaira
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV
; FILE REFERENCE: MBI-0025
; CURRENT APPLICATION NUMBER: US/09/934,455
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/227439
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: MBI-0022
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: MBI-0023
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 516
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 687
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-934-455-18

Query Match      27.9%; Score 852; DB 11; Length 687;
Best Local Similarity 37.1%; Pred. No. 1.5e-52;
Matches 229; Conservative 63; Mismatches 146; Indels 180; Gaps 24;

; 47 QGGLSE-----RTGSGVPKFKSTPPPSLPSPSPSSYFAI-----PGLSP 90
; 72 RGLSERIAARAGFNAPR-----LNTENIRTTDFSDSNLRSPCLTITSSPCLSP 121
; 91 AELDPSVLLNSNIIPLSPPTGAFVAQSFNWKSSSGGN--OOIVKEE--DKSFSNFSQ 145
; 122 ATLESFVFLNSPLAQPSPITG-----KFFFLPGVNGNALSSKAKDEPDDIGASFSEH 176
; 146 TRSGPPASSTATYQSSNVTVTQQPWMSFOEATKQDNFSSGKMMKTENSSMSQSFPEIA 205
; 177 ----PVSRSSSF-----FQGTTEMSVDYGNYNRSSHSHQSAEVPKGE 218
; 206 SVQTNHNSGQSDYGNYPQSQTL-----SRSD-----DG 237
; 219 NIESNLYGIETD--NQNGQNTSDVTTNTSLTVDHQEEEEQRRGDSMAGAPADGY 276
; 238 NWRKYGQKQVKGSENPSPRYKCTYPCNPTKKKVERSLDGOITEIVYKGTNNHPKQNTNR 297
; 277 NWRKYGQKLVKSEYPRSYKCTNPNCQVKKVEREGHITETIYKGAHNLKPPNRR 336
; 298 N-----SSNSSSLAIPHNSIRTEIPDQSYAT 324
; 337 SGQVQDGTGEVQEQOQOORSAATWVSCNNTQQCGGSNNV---EGSRTFEGNQSGL 393
; 325 HG--SQQMS-----AATPENSISIGDDDFEQQSQKCKS-----GGDEYDEDEDAKR 371
; 394 QAQTGGQYSGDGPVVVVVDASSTFSDNDEDDDRGTHGSVSLGYDGGGGGGGGESESKR 453
; 372 WKIEGENGMSAPGSTVREPRVVVQTTSDIDLDGQYRWKYGQKVKGNPNRSPRYK 431
; 454 RKLEAFAAEMSG--STRAIREPRVVVQTTSDVDILDDGQYRWKYGQKVKGNPNRSPRYK 512
; 432 THPGCPVRKHVERASHDLRAVITYYEGKNNHDPVAAAR-----GSGSHSVNRPMNNASNHT 487
; 513 TAPGCTVRKHVERASHDLKSVITYYEGKNNHDPVAAARNSHGGCGDSGN-----GNSG 565
; 488 NTAATSVRLLPVTHQSDNSLQONORSQAPPEG-----QSPFTLEMLQSPGSP----- 533
; 566 GSAAVS-----HHVYHNGHSEPPRGRFDROVTTNNQSPFS-----RPFSPQPHL 609
```


Db 95 SVGGGGSGDVPDFRFRQSRPTGLMITQPP-----GMFTVPPGLSPATLLDSPFFG---- 146
QY 106 LPSPPTTCFAVAQSFNWKSSGGNQIIVKEBDKSFNSFOTRSGPPASSTA-TYQSSNVT 164
Db 147 LFPPLQTF-----GMTHQALAQVTAQAVQGNVH 177
QY 165 VQTQPMFSQBATKQDNFSGKGMKMTENSSMQSFSPETASVQ-TNHSNGFQSDYQ--N 221
Db 178 MQSQSQSEYSSSTQCCQCCQCCQASLTEIPSFSSAPRSQIRASVQETSQGGQRETSEISVPE 237
QY 222 YPQSQTLRSRSDGYRWRYGQKVGKSNPRSYKCTYPNCPETKKKVERSLLDQITEI 281
Db 238 HRSQPNQADKPADDDGYRWRYGQKVGKSNPRSYKCTYPNCPETKKKVERSLLDQITEI 297
QY 282 VYKGNHHPQPNTRRNSNSSSLAI PHSNSIRTEIPDQSVATHGSGQMDSAATPENSSI 341
Db 298 IYKQGNHHELPQKGNNGSCKS-----SDIANQ-FQTSNSLNKSKDDQETSQV 346
QY 342 SIGDDDFEQSSQK-----CKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRVREPRVVQ 397
Db 347 TTTEQMSSEASDSERVGNAETSVGERHEDEPDPKRNTSEVRVSEPVASSHRTVTPEPRIIVQ 406
QY 398 TTSDIDLDGYRWRYGQKVGKSNPRSYKCTYPNCPETKKKVERSLLDQITEI 457
Db 407 TTSEVDLLDGYRWRYGQKVGKSNPRSYKCTYPNCPETKKKVERSLLDQITEI 466
QY 458 GKNHNDVPAARGSGSHSVNRPMPNPNASNTHTNT 489
Db 467 GKNHNDVPAARTS-SHQLR---PNNQHN-TST 493

RESULT 15
US-10-259-165-64
; Sequence 64, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; PRIOR FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 64
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-165-64

Query Match 24.7%; Score 753; DB 12; Length 266;
Best Local Similarity 51.3%; Pred. No. 5.2e-46;
Matches 162; Conservative 88; Mismatches 88; Indels 58; Gaps 8;
QY 267 KKKVERSL-DQQITEIYKGTTHNHPKQNTNRNSNSSSLAIPHNSIRTEIPDQSYATH 325
||||||| |||:||||| ||||| :|||||: |||

Db 2 KKKVERSLADGRITQIVYKGAHNHPKPLSTRNASSCATAA-----ACADDLAAPG 52
QY 326 GSGQMDSAATPENSSISIGDDDFEQSSQKCKSGGDEYDEDEPDAKRWKIEGENEGMS--- 382
Db 53 AGADQYSAATPENSSVTFGDDDEADNASHSEG-----DEPEAKEWKEDADNEGSSGGM 105
QY 383 --APGSRVREPRVVQTTSDIDLDGYRWRYGQKVGKSNPRSYKCTYPNCPETKKKVERSLLDQITEI 440
Db 106 GGGAGGKPVREPRLVVQTLSDIDLDGYRWRYGQKVGKSNPRSYKCTYPNCPETKKKVERSLLDQITEI 165
QY 441 HVERASHDLRAVITTYEGKHNHDVPAARGSGSHSVNRPMPNPNASNTHTNTAATSVRLLPVI 500
Db 166 HVERASHDTRAVITTYEGKHNHDVPAARGSGSHSVNRPMPNPNASNTHTNTAATSVRLLPVI 500
QY 501 HOSDNLQNRQSQAPPEQSGSPFTLEMLQSP-GSFGSGFGNPMQSYVNOQQQLSDNVSFSSR 559
Db 214 -----SAAAAQQQPYTLEMLPNPAGLYGGYGAG-----AGGAAFP 250
QY 560 TKEEPRDDMPLESLLC 575
Db 251 TKDERRDDLFVESLLC 266

Search completed: January 20, 2004, 16:25:08
Job time : 40 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:54 ; Search time 20 Seconds
(without alignments)
2764.850 Million cell updates/sec

Title: US-09-890-811B-10
Perfect score: 3050
Sequence: 1 MASSSGSLDTSASANSFTNF.....FSSRTKEPRDDMFLESLLC 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1822	59.7	549	2 S51529	SPF1 protein - swe
2	1331	43.6	514	2 S72443	DNA-binding protei
3	1139.5	37.4	512	2 T02498	probable WRKY-type
4	1032.5	33.9	402	2 S61413	DNA-binding protei
5	906	29.7	393	2 T00575	probable WRKY-type
6	888	29.1	571	2 T08930	hypothetical prote
7	884	29.0	309	2 T48481	SPF1-like protein
8	848.5	27.8	513	2 C84447	probable WRKY DNA-
9	840	27.5	528	2 T52092	DNA-binding protei
10	794.5	26.0	428	2 G86272	hypothetical prote
11	778.5	25.5	509	2 JC6203	sp8 binding protei
12	742	24.3	568	2 T05060	hypothetical prote
13	565.5	18.5	296	2 S72444	DNA-binding protei
14	561.5	18.4	349	2 E84790	probable WRKY-type
15	556.5	18.2	487	2 F84462	transcription fact
16	520	17.0	1895	2 T06609	disease resistance
17	488.5	16.0	782	2 B85362	hypothetical prote
18	316.5	10.4	528	2 C85056	probable DNA-bindi
19	306	10.0	548	2 T49348	hypothetical prote
20	301	9.9	327	2 T04919	DNA-binding protei
21	300	9.8	337	2 A84913	probable WRKY-type
22	299.5	9.8	458	2 T49114	hypothetical prote
23	295.5	9.7	252	2 B86422	FIN18.10 protein -
24	292.5	9.6	304	2 T09357	hypothetical prote
25	291.5	9.6	287	2 B96717	unknown protein, 3
26	288	9.4	294	2 T48026	hypothetical prote
27	288	9.4	471	2 B86322	hypothetical prote
28	275	9.0	374	2 G96704	probable DNA bindi
29	272.5	8.9	317	2 T00500	probable elicitor

30	264.5	8.7	489	2 B85022	probable DNA-bindi
31	259	8.5	349	2 S61414	DNA-binding protei
32	258.5	8.5	403	2 T02003	probable DNA-bindi
33	257.5	8.4	410	2 C84638	probable WRKY-type
34	257	8.4	341	2 S72445	DNA-binding protei
35	257	8.4	353	2 T09887	DNA-binding protei
36	251.5	8.2	324	2 T10685	hypothetical prote
37	245.5	8.0	380	2 C84710	hypothetical prote
38	235.5	7.7	302	2 A96841	hypothetical prote
39	235.5	7.7	387	2 C96720	hypothetical prote
40	235	7.7	97	2 A84899	probable WRKY-type
41	228.5	7.5	624	2 T05090	hypothetical prote
42	227	7.4	304	2 T05587	hypothetical prote
43	220.5	7.2	197	2 E84606	probable WRKY-type
44	220.5	7.2	421	2 H86431	T518.10 protein -
45	220.5	7.2	427	2 T00465	probable WRKY-type

ALIGNMENTS

RESULT 1

S51529

SPF1 protein - sweet potato

C;Species: Ipomoea batatas (sweet potato)

C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000

C;Accession: S51529

R;Ishiguro, S.; Nakamura, K.

Mol. Gen. Genet. 244, 563-571, 1994

O.

A;Title: Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that re-

A;Reference number: S51529; MUID:95058910; PMID:7969025

A;Accession: S51529

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-549 <158>

A;Cross-references: EMBL:D30038; NID:g484260; PIDN:BAA06278.1; PID:g484261

Query Match 59.7%; Score 1822; DB 2; Length 549;

Best Local Similarity 62.0%; Pred. No. 7e-101;

Matches 369; Conservative 66; Mismatches 92; Indels 68; Gaps 14;

QY	1	MASSSGSLDTSASANSFTNF	TFTSTHPTFTSFDLLASPLDNNKPPQGGI	-----SBR	TGS 56
DB	1	MAASGTIDAPTASSS	---FSFSTASSFMSFTDLLASDAYSGSVSRGLGDIARTGS	57	
QY	57	GVPKFKSTPPSLPLSPPI	SPSSYPAPPGLSPAELDSPVLLNSNLPSTTGAFVA	116	
DB	58	GVPKFKSLPPSLPLSSPA	VPSPSPYAPPGLSPELSDSPVLLSSNLPSTTGTPA	117	
QY	117	QSFNWKSSGGNQOI	IVKEEDKSFNSFOTRSGPPASSTATYQSSNVTVTQCPWSPQEA	176	
DB	118	QTFNWKNDNSAQOED	VKEEGYDPFSPQTNSS---ASMTLNVEDSK	160	
QY	177	TKQDNFSSGKGMKMT	TENSSMQSPFSPETASVQTNHNSN--GFQSDYQNY--PPOSQTL--SR	231	
DB	161	-----	-----RKDELNSLQSLPVTSTQWSSQNGSVSEYNNQCCPQTLREQR	207	
QY	232	RSDGYNWRKYQKQK	QVKGSENPSSYKCTYKCTPKKKKVERSLDGOITETVYKGTNNHPK	291	
DB	208	RSDGYNWRKYQKQK	QVKGSENPSSYKCTHPNCPTKKKVERALDGOITETVYKGAHHPK	267	
QY	292	PQNTRENS---	SNSSSLAI PHNSIRTEI PQOSVATHSGQMDSAATPENSISIGDDDD	347	
DB	268	POSTRSSSTASSAST	LAAQSYNAPSDVPQOSYWSNGQMDSVATPENSISVGGDDE	327	
QY	348	FEQSSCKSGGDEYDE	DEPDFDAKRWKIEGENGMSAPGSRTRVREPRVVQTTSDIIDDD	407	
DB	328	FEQSSCKSGGDEYDE	PDFDAKRWKIEGENGMSAPGSRTRVREPRVVQTTSDIIDDD	387	
QY	408	GYRWRYGQKQVKG	NPSPSSYKCTHPGCPVRKSHVERASHDLRAVITTYEGKHNHVPAA	467	
DB	388	GYRWRYGQKQVKG	NPSPSSYKCTSGQCPVRKSHVERASHDLRAVITTYEGKHNHVPAA	447	

QY 468 RGSQSHSVNPMNPASNHTNTAATSVRLLPVHQSDNSLQNG-----RQAPPEGO 519
 DB 448 RGSQSHGLNR-----GANPNNAAMVAIRP-----STMSQSNVPPIPIPTRMQOGEQ 498

QY 520 SPFTLEMLQSPGFGFGNPMQSYVNOQQLSNVFSSTKEEPRDDMFLESLL 574
 DB 499 APY--EMLQSGGFGYSGGNPMAYAN--QIQDNAP-SRAKEPRDDLFLDTLL 548

RESULT 2

S72443
 DNA-binding protein WRK1 - parsley
 C:Species: Petroselinum crispum (parsley)
 C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
 C/Accession: S72443
 R/Rushton, P.J.; Tovar Torres, J.; Parniske, M.; Wernert, P.; Hahlbrock, K.; Somssich, I.
 EMBO J. 15, 5690-5700, 1996
 A:Title: Interaction of elicitor-induced DNA-binding proteins with elicitor response ele
 A:Reference number: S72443; MUID:97051827; PMID:8836462
 A/Accession: S72443
 A:Molecule type: mRNA
 A/Residues: 1-514 <RUS>
 A/Cross-references: EMBL:U48831; NID:g1431871; PIDN:AAC49527.1; PID:g1431872
 C:Keywords: DNA binding

Query Match 43.6%; Score 1331; DB 2; Length 514;
 Best Local Similarity 52.5%; Pred. No. 9.2e-72;
 Matches 296; Conservative 72; Mismatches 128; Indels 68; Gaps 22;

QY 28 MTSFSLASPLDNNKPPQGLSERTGSGVVPKSTPPSLPLSPSPSPSVFALPPG 87

DB 1 MSSSLGDLAQPNDDFGSNWGFNQ-----KTKSPANQSLFPSPPPVPSFYFS----- 50

QY 88 LSPAEALDPSVLLNSSNLPSTTGAFAVQSFNWKSSGGNQIVKEEDKSFNFSQTR 147

DB 51 -----FLDSPIQNNNYIVSSGNGTWAQSF-----KEENQNFSDFSPPAQ 92

QY 148 SGPPASSTATVQSSNVTV-----QTQPPWSFOEATKQDNFSSGKGMKNTENSSMSQFS 201

DB 93 TRPASSTSSSLIPANTLVEDSLRKQKGNWFEEPAKKNDFLMDNASV-TSDIATLQRI 151

QY 202 PELASVQTN-HSN-GFQSDYGVNPPQSOVL-----SRRSDGYNWKYQKQVKGSENPR 254

DB 152 PEMTMQANNQSNAAALQSNLNVAQSSQSTNRDQSLDDGYNWKYQKQVKGSENPR 211

QY 255 SYKCTYPNCTKKKVERSLDQGITTEIVYKGTNNHHPQNTNRNNSNSSSLAIPHNSIR 314

DB 212 SYKCTYLCNCTKKKVETTFDGHITTEIVYKGNHHPKQSTKRSSQS-----YQNSIG 265

QY 315 TEIPDQSYATHGSGQMDSAATPENSISIGDDD-PEQSSQKCKSGGDEYDEPDARWK 373

DB 266 T-MPESLSLLE--NGRSEPVTTTPENSISLFGEDDLFEQGSNN-KPGDD--DGNBPDSCRWK 319

QY 374 IEGR-NEGMSAPGSRTRVREPRVVQTTSDIDLDGYNWKYQKQVKGPNPRSYKCT 432

DB 320 GEYESNPMESLGRTRVREPRVVQTTSDIDLDGYNWKYQKQVKGPNPRSYKCT 379

QY 433 HPGCFVRKHVERASHDLRAVITYYEGKHNHDVPAARGSGSH-SYNRPMNPNNASHHTTAA 491

DB 380 QVGCPRKHVERASHDLRAVITYYEGKHNHDVPAARGSGSVPAVNR-----SDNTTSAP 434

QY 492 TSVRLLPVHQSDNSLQNSQAPPEQSPFTLEMLQSPGFGSGGNPMQSY-VNQOQ 550

DB 435 TAIR--PTTNYL-NPLQNPRAQ-PANGQAPFTLEMLQSPGFGSGGNPMQSY-TNTYAINQ 490

QY 551 LSDNVFSSTKEEPRDDMFLESLL 574

DB 491 QASQCFST-AKDEPDVDSFDSFL 513

RESULT 3

T02498

probable WRKY-type DNA binding protein At2g38470 [imported] - Arabidopsis thaliana
 N:Alternate names: hypothetical protein T19C21.4
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C/Accession: T02498; D84805
 R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau
 submitted to the EMBL Data Library, August 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.
 A:Reference number: Z14676
 A/Accession: T02498
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A/Residues: 1-512 <ROU>
 A/Cross-references: EMBL:AC004683; NID:g3395421; PID:g3395425
 A:Experimental source: cultivar Columbia
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A/Accession: D84805
 A:Status: preliminary
 A:Molecule type: DNA
 A/Residues: 1-512 <STO>
 A/Cross-references: GB:AE002093; NID:g6598471; PIDN:AAC67339.2; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g38470; T19C21.4
 A:Map position: 2
 A:Introns: 74/3; 143/3; 321/2; 375/2

Query Match 37.4%; Score 1139.5; DB 2; Length 512;
 Best Local Similarity 44.7%; Pred. No. 2.2e-60;
 Matches 254; Conservative 80; Mismatches 144; Indels 90; Gaps 15;

QY 40 LDNNKPPQ-----GGLSERTGS-----GVPKFKSTPPSLPLSPSPSPSVFALPP 85

DB 1 MDNSTRQNMNGSANWSQQSGRTSTSSLEDLEIPKPSFAPSISISPSLSVSPSTCF-- 57

QY 86 PGLSPAEALDPSVLLNSS-NILPSPPTGAFVAQSFNWKSSGGNQIVKEEDKSFN-- 141

DB 58 ---SPSLFLDSPAFVSSSANVLASPTTGALI-----TNVTNKGINEGDKSNNNFN 106

QY 142 ---FSFOTRSGPPASSTATVQSSNVTVQTPQPSFOEATKQDNFSSGKGMKNTENSSMSQ 198

DB 107 LDFDFSHQSGSVSAFTTTTTTTTTTNS--SIFSQEQKKQSEQWSQTETRPNQ 164

QY 199 SFSPEIASVQTNHNSGFSQSDYGVNPPQSOQLSRRSDGYNWKYQKQVKGSENPRSYK 258

DB 165 AVS-----YNGRE-----QRKGEDGYNWKYQKQVKGSENPRSYK 201

QY 259 CTYPNCTKKKVERSLDQGITTEIVYKGTNNHHPQNTNRNNSNSSSL-----AIPHNS 311

DB 202 CTYPNCTKKKVERSLDQGITTEIVYKGTNNHHPQNTNRNNSNSSSLFHSAYVNASLDHNR 261

QY 312 SIRTEIPDQSYATHGSGQMDSAATPENSISIGDDDPEQSSQKCKSGGDEYDEPDAR 371

DB 262 QASDQPNNSNPHQSDSFGMQQEDNTTSDSVGDDFEQGS-SIVRDEDCGSEPEAKR 320

QY 372 WKTEGENEGMSAPGSRTRVREPRVVQTTSDIDLDGYNWKYQKQVKGPNPRSYK 431

DB 321 WKGDNETNGNGGSGKTVREPRIVVQTTSDIDLDGYNWKYQKQVKGPNPRSYK 380

QY 432 THGCCPRKHVERASHDLRAVITYYEGKHNHDVPAARGSGSHSVNPMNPNNASHHTTAA 491

DB 381 TTIGCFVRKHVERASHDLRAVITYYEGKHNHDVPAARGSG-YATNRAPOQSSSVFTRPAA 439

QY 492 TSVRLLPVHQSDNSLQNSQAPPEQSPFTLEMLQ-----SPGSGFGSGGNPMQSYVN 547

DB 440 IA-----GHSNYTSSQAPYTLQMLHNNNTNTGTFPGYAMNNNNNSNLQ 483

QY 548 QQQLSDNVFSSTKEEPRDD-MFLESLL 574

Qy 342 SIGDDDFEQSQK-----CKSGGDYDDEPPDAKRWKIENEGNSAPGSRIVREPRVVVQ 397
Db :
347 TTTEQMSEASDSVEVGNAETSVGERHEDEPDPKRNTETVRVSBPVASSHRTVTTEPRIIVQ 406
Qy 398 TTSIDILDDGYRKWKYGQVKVGNPNPRSYYKCTHPGCCPVKRKHVERASHDLRAVITTYE 457
Db :
407 TTSEVDLDDGYRKWKYGQVKVGNPNPRSYYKCTTPDCGVKRKHVERAATDPKAVITTYE 466
Qy 458 GKNHNDVPAARGSGSHSVNRPMPNNASNHTNT 489
Db :
467 GKNHNDVPAARTS-SQLR---PNNQH-NST 493

RESULT 9
T52092
DNA-binding protein WRKY2 [imported] - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T52092
R;Wang, Z.; Yang, P.; Pan, B.; Chen, Z.
Plant J. 16, 515-522, 1998
A:Title: An oligo selection procedure for identification of sequence-specific DNA-binding proteins
A:Reference number: Z25947; MUID:99097845; PMID:9881170
A:Accession: T52092
A>Status: preliminary; translated from GB/EMBL/DDBB
A:Molecule type: mRNA
A:Residues: 1-528 <WAN>
A:Cross-references: EMBL:AF096299; PIDN:AAD16139.1
C:Genetics:
A:Gene: WRKY2

Query Match 27.5%; Score 840; DB 2; Length 528;
Best Local Similarity 38.7%; Pred. No. 1.4e-42;
Matches 206; Conservative 68; Mismatches 142; Indels 116; Gaps 17;

Qy 15 NSFTNFTHPTFPMTTFSDLLASPLDN-----NKPPGGLS-----ERTSGVP 59
Db :
45 SSFSDPASEYP----SFSQLAGAMSPAAFSQORGFPPPPPMTVISIKGESGSDM 100
Qy 60 K--FKSTPPPSLPSPPISSFYAIIPGLSPAELLDSPVLLNSSNILPSTTGAFVAQ 117
Db :
101 DFGFKQRPSGLVITQSEM----FTIPAGLSPARLLGSLPF-----SPQGPGGMS 148
Qy 118 SPNWKSSGGNQIIVKEDKSFNFSQTSGPPASATAYQS--SNVTQTQPWSFOE 175
Db :
149 --HOQAQVTAQAAPHQSQMHIOPTYDPSASAAPAFSOFQSLTSNATANKQIPPP--- 203
Qy 176 ATKQDNFSSGGMKMTENSMSQSFSEIASVQTNHNGFQSDYGNYPPOSQTLSSRDD 235
Db :
204 -----ASDPNVMEASEVSLDQSE-----PASSAVDKPAD 236
Qy 236 GYNWRKYGQKVKGSENPRSYKCTYPCNCPTKKKVERSLEDGQITEIVYKGTNHHPKPNT 295
Db :
237 GYNWRKYGQKHVGSEYPRSYKCTHPENCVKKVERSLEDGQVTEIIYKGNHQHPQSS 296
Qy 296 RNSNSNSSLAI PHSNSIRTEI PQOSVATHSGQOMDS---AATPENS-----STSIGD 345
Db :
297 KRKSEGN-----PNGNILOQPSLESEGVA GTLNNSKDSMPYSLRMTD 342
Qy 346 DDFEQSQKCKSGGDEYDE-----DEPAKWKEGE-NEGMSAPGSRGTVREPR 393
Db :
343 QESSQATHDVSGTISEGEEVGD TENADGNDERESKERAIEVQTSEAA SASHRAVAEP 402
Qy 394 VVOTTSDIDLDDGYRWRYKYGQVKVGNPNPRSYYKCTHPGCCPVKRKHVERASHDLRAVI 453
Db :
403 IIVQTTSEVDLDDGYRWRYKYGQVKVGNPNPRSYYKCTSGCNVNRKHVERASPDKAVI 462
Qy 454 TTYEGKNHDVPAARGSGSHSVNRPMPNNASNHTNTAATSVR--LLPVIHOSD 504
Db :
463 TTYEGHNHDVFAAR-----NSHNTNNSVSQMRPHNPVVDKOD 502

Qy 342 SIGDDDFEQSSQK-----CKSGGDYDDEDPDAKKWKIEGNEGMSAPGSRTVREPRVVQ 397
Db :
347 TTTEQMSEASDSVEGNAETSVGERHEDPDPKRNTFVRVSEFPVASSHRTVTTEPRIIVQ 406
Qy 398 TTSIDIDLDDGYRWKYGQVKVGKNPNPRSYYKCTHPGCCPVKRKHVERASHDLRAVITTYE 457
Db :
407 TTSEVDLDDGYRWKYGQVKVGKNPYRSYYKCTTTPDCGVKRKHVERAAATDKPAVITTYE 466
Qy 458 GKNHNDVPAAAGSGSHSVNRPMPNNASNHTNT 489
Db :
467 GKNHNDVPAARTS-SHQLR---PNNQH-NST 493

RESULT 9
T52092
DNA-binding protein WRKY2 [imported] - common tobacco
C/Species: Nicotiana tabacum (common tobacco)
C/Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C/Accession: T52092
R/Wang, Z.; Yang, P.; Fan, B.; Chen, Z.
Plant J. 16, 515-522, 1998
A>Title: An oligo selection procedure for identification of sequence-specific DNA-binding proteins
A/Reference number: 225947; MUID:99097845; PMID:9881170
A/Accession: T52092
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-528 <WAN>
A/Cross-references: EMBL:AF096299; PIDN:AAD16139.1
C/Genetics:
A/Gene: WRKY2

Query Match 27.5%; Score 840; DB 2; Length 528;
Best Local Similarity 38.7%; Pred. No. 1.4e-42;
Matches 206; Conservative 68; Mismatches 142; Indels 116; Gaps 17;

Qy 15 NSFTNFTHFTSPFWMTTFSOLLASPLDN-----NKPPGGGLS----ERTSGVGP 59
Db :
45 SSFTSDPASEYP---SFSQLLAGAMASPAAFSGQRGFPPPEMTVSISKGESGSDM 100
Qy 60 K--EKSTPTPSLPSPPISSSYFAIPPGLSPAELLDSPVLNSSNILSPPTGAFAVAQ 117
Db :
101 DFGFKQNRPSGLVITQSPM-----FTIPAGLSPARLLGSLFF-----SPQGPGFMS 148
Qy 118 SFNMSSSGGNQIWKSEDKSFNSFOTRSGPPASSTATVQS--SNVTQTQOPWSFQE 175
Db :
149 --HQOALQAQTAQAAHPQSQWHIQDPYSSSAAPAPSFOQLTSNATANKQIPP--- 203
Qy 176 ATKODNFSSGKMKTNSMSQSFSPFIASVQTNHNGFQSDYGVNPPOSQTLRRSDD 235
Db :
204 -----ASDPNVMEASEVSLSDORSE-----PASSAVDKPAD 236
Qy 236 GYNWRKYGQKVKGSENPERSYYKTYPCNCPKTKKVERSLEDGOITEIVYKGTNHHPKPNT 295
Db :
237 GYNWRKYGQKVKGSENPERSYYKTHPCNCPKTKKVERSLEDGOITEIVYKGTNHHPKPNT 296
Qy 296 RNSSNSSSLAI PHSNSIRTEIPQSYATHSGSQWMDS--AATPENS-----SISIGD 345
Db :
297 KRKSKEGN-----PNGNTNLQPSLSSEGAVGTLNNSKMSNPYSURMTD 342
Qy 346 DDFEQSSQKCGGDEYDE-----DEPAKWKIEGE-NEGMSPAGSRTRVREPR 393
Db :
343 QESSQATHDVSGTSGEEVGD TENADGNDERESKERRAEIVQTSEASSASHRAVAEP 402
Qy 394 VVOTTSDIDLDDGYRWKYGQVKVGKNPNPRSYYKCTHPGCCPVKRKHVERASHDLRAVI 453
Db :
403 IIVOTTSEVDLDDGYRWKYGQVKVGKNPYRSYYKCTSGCNVCRKHVERAPSDPKAVI 462
Qy 454 TTYEGKNHNDVPAARGSGSHSVNRPMPNNASNHTNTAATSVR--LLPVIHQSD 504
Db :
463 TTYEGHNHDVPAAR-----NSHNTTNNVSQMRPNHPVVDKOD 502

```
Db      TTYEGEHHNDVPAAR-----NSSHTTNNNSQMRPHNPVVVDKQD 502
```

C;Species: Cucumis sativus (cucumber)
 C;Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 21-Jul-2000
 C;Accession: JC6203
 R;Kim, D.J.; Smith, S.M.; Leaver, C.J.
 Gene 185, 265-269, 1997
 A;Title: A cDNA encoding a putative SPFI-type DNA-binding protein from cucumber.
 A;Reference number: JC6203; MUID:97208883; PMID:9055825
 A;Accession: JC6203
 A;Molecule type: mRNA
 A;Residues: 1-509 <KIM>
 A;Cross-references: GB:141134; NID:G927024; PID:AAC37515.1; PID:G927025
 C;Comment: This protein is a DNA-binding protein consisting of a monomer. It is involved
 C;Genetics:
 A;Gene: sz71
 C;Keywords: DNA binding

Query Match 25.5%; Score 778.5; DB 2; Length 509;
 Best Local Similarity 39.9%; Pred. No. 5.9e-39;
 Matches 196; Conservative 63; Mismatches 167; Indels 65; Gaps 15;

QY 31 SFDLLASPLDNKPPGGISERTGSGVPKFKSTPPSLPL-----SPP---PISPSSY 81
 DB 61 SFSQLLAGAMAS---PMAMGFFGTGS-TPNYAKDGGPASELEFGMKQSKPVLNLVARSPL 116
 QY 82 FATPPGLSPAELLDSPVLLNSSNLTPTTGAFVAQSFNWKSSGGNQQIVKEEDKSFN 141
 DB 117 FSVPPGLSPGLLNSP-----GFYPPQSPFGMSHQQAALAQVTAQAALANSH 162
 QY 142 FSFQTRSGPPASPTATYQSSNVTVTQTPWSPQEQATKQDNFSGKGMKMTENSSMQSFS 201
 DB 163 MHMQ-----QAEYOHSSVPAPT-EPLVRDPSFSLDASQALAIPTSDTKSLIAES 212
 QY 202 PETASVQTNHNS-GFQSDYGNYPQSQTLRRSDGYNWRKYQKQKVGKSENPRSYKCT 260
 DB 213 TEV-----SHSDRYQP-----PPPPRGSKDPAADDGYNWRKYQKQKVGSEFPRSYKCT 262
 QY 261 YPNCPTKKKVERSLDGQITETVYKGTNHPKQPNTRNSNSSSLAIPHNSNIRTEIPDQ 320
 DB 263 HLNCPRKKKIEGLPDGTEITLIYKQNNHPPPPANKARDNIEPAGCTNS-LIKPEGCLQ 321
 QY 321 SYATHGSGQMDSAATPENSISIGDDDFEQQSKCKSGGDEYDEDEPDAKRWKIEGENEG 380
 DB 322 NQA--GILNKSENVLGSSDSEGRADTEIT-----DDRDEDEPNKRQINDAGTSG 371
 QY 381 MSAPGSTRVREPRVVQTTSDIDLDGYNWRKYQKQKVGKSENPRSYKCTHPGCPVRK 440
 DB 372 V-ALSHKTLTEPKIIIVQTRSEVDLLDDGYRWKRYQKQKVGKSENPRSYKCTSAGCNVRK 430
 QY 441 HVERASHDLRAVITTYEGKHNHVPAAARGSGSHSVNRPMPNPNASNHTNTAATSVRLLPVI 500
 DB 431 HVERSSDTSKAVVTTYEGKHNHVPAAARNSSHHTV-----NNTVHHIKPLKVVAQKPILL 485
 QY 501 HQSDNSLQNR 511
 DB 486 KEMEFGTNDQR 496

RESULT 12
 T05060
 hypothetical protein M3E9.130 - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C;Accession: T05060
 R;Bevan, M.; Vandenbol, M.; Jallet, C.; Portetelle, D.; Hoheisel, J.; Mewes, H.W.; Mayer
 submitted to the Protein Sequence Database, March 1999
 A;Reference number: Z15396
 A;Accession: T05060
 A;Molecule type: DNA
 A;Residues: 1-568 <BEV>
 A;Cross-references: EMBL:AL022223
 A;Experimental source: cultivar Columbia; BAC clone M3E9
 C;Genetics:
 A;Map position: 4

A;Introns: 106/3; 339/2; 392/2
 A;Note: M3E9.130

Query Match 24.3%; Score 742; DB 2; Length 568;
 Best Local Similarity 37.3%; Pred. No. 9.9e-37;
 Matches 193; Conservative 55; Mismatches 119; Indels 150; Gaps 16;

QY 86 PGLSPAELLDSPVLLNSSNLTPTTGAFVAQSFNWKSSGGNQQIVKEEDKSFNFSQ 145
 DB 88 PGLSPATLLESFVFL--SNPLLSPTTKGL-----SSVPSDKAKAELFDDITTSAPQ 137
 QY 146 TRSGPPASPTATYQSSNVTVTQTPWSPQEQATKQDNFSGKGMKMTENSSMQSPEA 205
 DB 138 TISGGLDPT-----NIALEPDDSDQYEERQ-----LGGLG-----DSMACCAP--- 176
 QY 206 SVQTNHNSGFSQDYGNYPPQSQTLSRRSDGYNWRKYQKQKVGKSENPRSYKCTVPNCP 265
 DB 177 -----ADDGYNWRKYQKQKVGSEYPRSYKCTHPNCE 209
 QY 266 TKKKVERSLDGQITETVYKGTNHPKQPNTRNSNSSSLA----- 306
 DB 210 AKKKVERSREGHILEIITGDHHSKPPNRRSGISGSGTGQDMQIDAIEYEGFACTNEN 269
 QY 307 IPHNSIRTEIPDQSYA-----THGSGQMDSAATPENSISIGDDDFEQQSKCKSG 358
 DB 270 IEWTSFVSAELEYGSHSGSMQVQNGTHQFGYDAAA-----DALYRDENEEDRTSHMSVSL 325
 QY 359 GDEYDEDEPDAKRWKIEG--ENEGMSAPGSTRVREPRVVQTTSDIDLDGYNWRKYG 415
 DB 326 TYDGEVEESESRRKLEAYATETSG-----STRASREPRVVQTTSDIDLDGYNWRKYG 381
 QY 416 QKVKGKGNPNRPSYKCTHPGCPVRKHVERASHDLRAVITTYEGKHNHVPAAAR-----GS 470
 DB 382 QKVKGKGNPNRPSYKCTANGCTVTKHVERASDDFKSVLTYYIGKHTVVVPAARNSSHVGA 441
 QY 471 GS-----HSVNRMPNPNASNHTNTAATSVRLLPVIHQSDNSLQNRQAAP 516
 DB 442 GSSGTLQGLSATQTNHNVHVPMPHSRSEGLATA-----NSSLPDFQSH--- 485
 QY 517 EGQSPFTLEMLQSPGSGFGFGNPMQSVVNOQLSD 553
 DB 486 -----LRHPTGFSV-----YIGQSELSD 503

RESULT 13
 S72444
 DNA-binding protein WRKY2 - parsley (fragment)
 C;Species: Petroselinum crispum (parsley)
 C;Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
 C;Accession: S72444
 R;Rushton, P.J.; Tovar Torres, J.; Parniske, M.; Wernert, P.; Hahlbrock, K.; Somssich,
 EMBO J. 15, 5690-5700, 1996
 A;Title: Interaction of elicitor-induced DNA-binding proteins with elicitor response el
 A;Reference number: S72443; MUID:97051827; PMID:8896462
 A;Accession: S72444
 A;Molecule type: mRNA
 A;Residues: 1-296 <RUS>
 A;Cross-references: EMBL:U58540; NID:G1432057; PIDN:AAC49529.1; PID:G1432058
 C;Keywords: DNA binding

Query Match 18.5%; Score 565.5; DB 2; Length 296;
 Best Local Similarity 43.0%; Pred. No. 1.3e-26;
 Matches 132; Conservative 36; Mismatches 68; Indels 71; Gaps 9;

QY 271 ERLDGOITETVYKGTNHPKQPNTR-----NSSSSSLAIPHNSIRTEIP 318
 DB 1 ERLDGOITETVYKGTNHPKQPNTRRFTAGALISIQEEKAVNASL-----TGQG 51
 QY 319 DQSYATHGSGQMDSAATPENSISIGDDDFEQQSKCKSGGDEYDEDEPDAKRWKIEGEN 378
 DB 52 DTTISQTLA--DQNGTPIUSSPRGVNADVNGASPLNSVTDDIDNDQDFMKRRRTDVG 109
 QY 379 EGMSPGSTRVREPRVVQTTSDIDLDGYNWRKYQKQKVGKGNPNRPSYKCTHPGCPV 438

Db 110 IDIS-PVVKIREPRVVVQTLSEVDILDDGYRWKYGQKVRGNPNRPRSYKCTNAGCPV 168
QY 439 RKVERASHDLRAVITTYEGKHNHVDPAARGSGSHSVNRMP-----480
Db 169 RKVERASHDPKAVITTYEGKHNHVDPAK-TNSHDVSGSAPISGSRVPEPCSPISLD 227
QY 481 -----NN-----ASNHTNT---AATSVRLLPVIH-----QSDNSLQNR 512
Db 228 LGVMGNYGVNNGVEQQTLLGSEHVRVQVASTAPTCTGYGVANGIILYGRDNHVENHF 287
QY 513 QAPPEQ 519
Db 288 ESPPLGQ 294

RESULT 14

E84790
probable WRKY-type DNA binding protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84790
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84790
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <STO>
A:Cross-references: GB:AE002093; NID:g4056481; PIDN:AAC98047.1; GSPDB:GN00139
A:Gene: At2g37260
A:Map position: 2

Query Match 18.4%; Score 561.5; DB 2; Length 349;
Best Local Similarity 37.1%; Pred. No. 2.8e-26;
Matches 140; Conservative 45; Mismatches 109; Indels 83; Gaps 11;
QY 134 BEKKSFPN-SFOTRSGPPASSTATYQSSNVTVTQOPWQFQATKQDNFSSGKGMKTE 192
Db 4 DDDSDSRNVVYKPKAKLVSKATVSAANLMQGNRQOTWRQSEAV-----SYGKSV--- 54
QY 193 NSSSMQSFSEIASVQTNHNGFSQSDYGNYPQSQTLRSDDGYRWKYGQKVGSEN 252
Db 55 -SQGTHRAGNLVQKVPSTES-----ETSTGDRSSVDGYRWKYGQKVGSGEC 103
QY 253 PRSYKCTYPCPTKKKVERSLDQIPEIVYKGTNNHPKQ--NTRNRSNSSSLAIPHS 310
Db 104 PRSYKCTHPKCPVKKVERSEGVSEIYVQGEHNSKSCPLPRASSISS-----157
QY 311 NSIRTEIPDOSYATHGSGQWD-----SAATP 336
Db 158 ---GFOKPKPSIASGSMQDPPNNLYSLPNNQSDSTQNRTEKMSGCVITPFFFAVP 214
QY 337 ENSSISIGDDDFEQSQCKSGGDEYDEPD-AKRWKIEGENEGHSGAPSRVTRPRVV 395
Db 215 RSTNSNPGTSDGCKSQSC---DEGLDDPFSKRRK---NEKOSSEAG-----V 258
QY 396 VQTTSDIIDLDDGYRWKYGQKVGKGNPNRPRSYKCTHPGCPVKRKHVERASHDLRAVIT 455
Db 259 SQGSVESDSDLEDGRWRKYQKVGNGAYPRSYRCTISANCRARKHVERASDDPRAFIT 318
QY 456 YEGKHNHVDPAARGSGS 472
Db 319 YEGKHNHLLSPSSS 335

RESULT 15

F84462
transcription factor ZAP1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84462
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84462
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-487 <STO>
A:Cross-references: GB:AE002093; NID:g4585919; PIDN:AAD25579.1; GSPDB:GN00139
A:Gene: At2g04880
A:Map position: 2

Query Match 18.2%; Score 556.5; DB 2; Length 487;
Best Local Similarity 35.0%; Pred. No. 8.5e-26;
Matches 131; Conservative 50; Mismatches 98; Indels 95; Gaps 10;
QY 234 DDGYRWKYGQKVGSENPRSYKCTYPCPTKKKVERSLDQIPEIVYKGTNNHPK- 292
Db 111 EDGYRWKYGQKLVKGNFVRSYRCTHPNCKAKQLERSAGQVVDVTFGEHDPKPL 170
QY 293 -----QNTRNRSNSSS-----LAI-----PHSNSIRTEIPDOS 321
Db 171 AGAVPINQDKRSDVFTAVSKGQRDIVSLIYKLCIVSYDIMFVEKTSGSSVQTLRQTEP 230
QY 322 YATHGSGQWDSAAATPENSISIGDDDFEQSQCKSGGDEYDEPD--EPDAKRWKIEGENE 379
Db 231 PKIHGGLHV-----SVIPPADDVKTDISQSSRITGDNTHKDVNSPTAKRRKKGNI 282
QY 380 GMSAPGSRVTRPRVVVQTTSDIIDLDDGYRWKYGQKVGKGNPNRPRSYKCTHPGCPVR 439
Db 283 --LSPVERSTNDSRIVVHTQTLLFDIVNDGYRWKYGQKSVKSGSPYRYSRCSGPCPVK 340
QY 440 KHEVERASHDLRAVITTYEGKHNHVDPAARGSGSHSVNRMPNPNASNHTNTAATSVLLPV 499
Db 341 KHEVERASHDTKLITTYEGKHDHDP-----FGRVVTHN-----375
QY 500 IHQSDNSLQNRSQAPPEGQSPFTLEMLQSPGSGFGNPMQSYVNOQQLSDNVFSSR 559
Db 376 --MLDSEVDDKEGDANKTPQS-----STLQSTTKQHVHEDHL-----410
QY 560 TKEEPRDDMFLES 573
Db 411 -RKTKTNGFEKSL 423

Search completed: January 20, 2004, 16:20:29
Job time : 23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:52 ; Search time 18 Seconds
(without alignments)
1502.242 Million cell updates/sec

Title: US-09-890-811B-10

Perfect score: 3050

Sequence: 1 MASSGSLTSASANSFTNF.....FSRTKEPRDDMFLESLLC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1139.5	37.4	512	1 WR33_ARATH	Q88p5 arabidopsis
2	906	29.7	393	1 WR25_ARATH	O22921 arabidopsis
3	899	29.5	557	1 WR20_ARATH	Q93w70 arabidopsis
4	884	29.0	309	1 WR26_ARATH	Q9c5t3 arabidopsis
5	861.5	28.2	514	1 WRK4_ARATH	Q9xi90 arabidopsis
6	852	27.9	687	1 WRK2_ARATH	Q9fg77 arabidopsis
7	848.5	27.8	513	1 WRK3_ARATH	Q9zq70 arabidopsis
8	742	24.3	568	1 WR34_ARATH	O65590 arabidopsis
9	738	24.2	423	1 WR58_ARATH	Q93wu7 arabidopsis
10	566.5	18.6	429	1 WR44_ARATH	Q9zuu0 arabidopsis
11	556.5	18.2	487	1 WRK1_ARATH	Q9ai37 arabidopsis
12	520	17.0	1895	1 WR19_ARATH	Q98z67 arabidopsis
13	482	15.8	466	1 WR32_ARATH	P59583 arabidopsis
14	386	12.7	506	1 WR10_ARATH	Q91g05 arabidopsis
15	348	11.4	399	1 WR48_ARATH	Q9fgz4 arabidopsis
16	318	10.4	218	1 WR12_ARATH	Q93wv4 arabidopsis
17	316.5	10.4	528	1 WR42_ARATH	Q9xec3 arabidopsis
18	315.5	10.3	318	1 WR28_ARATH	Q8vwj2 arabidopsis
19	310.5	10.2	553	1 WRK6_ARATH	Q9c5j9 arabidopsis
20	309.5	10.1	326	1 WRK8_ARATH	Q9f126 arabidopsis
21	306	10.0	548	1 WR72_ARATH	Q91xg8 arabidopsis
22	305.5	10.0	480	1 WR61_ARATH	Q8vwv6 arabidopsis
23	303.5	10.0	538	1 WR31_ARATH	Q93wv0 arabidopsis
24	301	9.9	282	1 WR71_ARATH	Q93wv4 arabidopsis
25	300	9.8	337	1 WR23_ARATH	O22900 arabidopsis
26	292.5	9.6	304	1 WR13_ARATH	Q9svb7 arabidopsis
27	291.5	9.6	287	1 WR78_ARATH	Q9c983 arabidopsis
28	288	9.4	277	1 WR68_ARATH	Q93wv6 arabidopsis
29	280.5	9.2	147	1 WR45_ARATH	Q98763 arabidopsis
30	275	9.0	374	1 WRK9_ARATH	Q9c9f0 arabidopsis
31	272.5	8.9	317	1 WR15_ARATH	O22176 arabidopsis
32	268	8.8	195	1 WR56_ARATH	Q8vwq4 arabidopsis
33	264.5	8.7	145	1 WR75_ARATH	Q9fy22 arabidopsis

34	264.5	8.7	489	1	WR47_ARATH	Q9zsi7 arabidopsis
35	259.5	8.5	179	1	WR24_ARATH	Q9ffs3 arabidopsis
36	257	8.4	353	1	WRK7_ARATH	Q9stx0 arabidopsis
37	256	8.4	173	1	WR50_ARATH	Q8vqv5 arabidopsis
38	251.5	8.2	324	1	WR11_ARATH	Q8vq15 arabidopsis
39	247	8.1	321	1	WR17_ARATH	Q9sja8 arabidopsis
40	245.5	8.0	380	1	WR21_ARATH	O04336 arabidopsis
41	243	8.0	194	1	WR51_ARATH	Q93wu9 arabidopsis
42	240	7.9	330	1	WR74_ARATH	Q93wv6 arabidopsis
43	236.5	7.8	330	1	WR39_ARATH	Q9s07 arabidopsis
44	236	7.7	109	1	WR43_ARATH	Q8gvl1 arabidopsis
45	235.5	7.7	302	1	WR40_ARATH	Q9sh7 arabidopsis

ALIGNMENTS

RESULT 1
WR33_ARATH STANDARD; PRT; 512 AA.
ID AC Q8S8P5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 33 (WRKY DNA-binding protein 33).
GN WRKY33 OR AT2G38470 OR T19C21.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RA Lippok B., Somssich I.E.;
RT "Arabidopsis thaliana transcription factor WRKY33."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RC MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ranning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Unayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana."
RL Nature 402:761-768(1999)
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the WRKY class I family.
CC -!- SIMILARITY: Contains 2 WRKY domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF509499; AAM34736.1; -;
EMBL; AC004683; AAM14994.1; -;
PIR; T02498; T02498.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.

Db 58 QSH-----NFTSDYLDSPILLSHSLSPITGTFPLQFGNCTN----- 99
Qy 138 SFNFQTSRGPASSTATYQSSNVTVQTPQWFSOEATKQDNFSGKGMKMTSSSM 197
Db 100 NHDSPQLSQP-----SNASSALQETVYGVQDHEK----- 131
Qy 198 QSFSP-EIASQVTHNSGFDYGVNYPQSQCTLSRRSDGYNWRKYGQKQVKGSENPSSY 256
Db 132 QEMIPNEIATONNQSFGTQRI-KIP--AYMVSNSNDGYNWRKYGQKQVKGSENPSSY 188
Qy 257 YKTCYPCPKKXKVERSLDGOITIVYKVTGTHNPKPQNTNRNSSSSLAIPHS----- 310
Db 189 FKTYPCVSKKIVETASQCTIETIYKGGHNPKEPTKRPQSS-----LPSSVNGRRL 244
Qy 311 ---NSIRTEPDOSYATHGSGQMSAATPNSISIGDDDFEQSOKCKSGGDEYDEDEP 367
Db 245 FNPASVSEPHDS-----ENSSISFDYSLEQKSPKSEGEIDEEEP 289
Qy 368 DAKRWKIEGNEGMSAPGSTVREPRVVVQTTSDIDILDGYRWKYGQKQVKGSENPSSY 427
Db 290 EMKRMKREGDEGMSIEVSKGKPEPRVVVQTTSDIDILDGYRWKYGQKQVKGSENPSSY 349
Qy 428 YKCTHGPCVRKHVERASHDLRAVITTYEGKNHDPAA 467
Db 350 YKCTFGCGVKQKQVKSAADEKRAVLTYEGRNHDIPTA 389

RESULT 3
WR20 ARATH STANDARD; Q94AP6; Q9SU40; PRT; 557 AA.
AC Q93WV0; Q8H1E9; Q94AP6; Q9SU40;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 20 (WRKY DNA-binding protein 20).
GN WRKY20 OR A74G2640 OR T15N24.90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 1).
RP STRAIN=cv. Columbia; TISSUE=Flower;
RC MEDLINE=20083488; PubMed=10617198;
RX Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boudry M., Bancroft I.,
RA Vos P., Hohseil J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirks W.,
RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Bernheiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysshaert C., Gielens J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Bense V., Rechmann S.,
RA Borkova D., Bloeker H., Scharf M., Grimm M., Loeferl T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Loeferl T.-H.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,

RA Massenot O., Quigley F., Clabaud G., Muendlein A., Feiber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse C.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stonking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vill D., Shekner M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana";
RT Nature 402:769-777(1999).
RN [3]
SEQUENCE FROM N.A. (ISOFORM 2).
RP STRAIN=cv. Columbia;
RC Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAPFs) sequenced by the
SSP consortium (Salk/Stanford/PGEC).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE FROM N.A. (ISOFORM 2).
RP STRAIN=cv. Columbia;
RC Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Arabidopsis open reading frame (ORF) clones";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q93WV0-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=2;
CC IsoId=Q93WV0-2; Sequence=VSP_007247, VSP_007248;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Belongs to the WRKY group I family.
CC -!- SIMILARITY: Contains 2 WRKY domains.
CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AF425837; AAL13050.1; -;
DR EMBL; AL078465; CAB43860.1; ALT_SEQ.
DR EMBL; AL161565; CAB79519.1; ALT_SEQ.
DR EMBL; AY045892; AAK76566.1; -;
DR EMBL; AY150436; AAN12978.1; -;
DR PIR; T08930; T08930.
DR InterPro; IPR003657; WRKY.

```
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS00811; WRKY; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;
FT Alternative splicing.
FT DNA_BIND 205 269 WRKY 1.
FT DNA_BIND 375 440 WRKY 2.
FT FT 1 72 Missing (in isoform 2).
FT VARSPLIC 1 72 /FTid=VSP_007247.
FT VARSPLIC 73 76 SNIK -> MILL (in isoform 2).
FT VARSPLIC 108 108 /FTid=VSP_007248.
FT CONFLICT 108 108 O -> R (IN REF. 3).
SQ SEQUENCE 557 AA; 61033 MW; 61793844AA4CEFD09 CRC64;

Query Match 29.58; Score 899; DB 1; Length 557;
Best Local Similarity 40.48; Pred. No. 7.6e-43;
Matches 226; Conservative 64; Mismatches 165; Indels 104; Gaps 20;

QY 55 GSGVPKFKSTPPSLPLSPPISSYFAIPPLGSLPAELDPSVLLNSNLSPTTGA- 113
DB 32 GGGARYKLMSPAKL-----PISRSTDTIPPGLSPTSPLESFVFNISNIKPEPSTGSL 86
QY 114 FVAQSFNWKSSG-----GNOQIVKEEDKFSNFSFQTRSGPPASSTATVQSSNVVQTQ 168
DB 87 FKPRPVHISASSSYTGRGFHNTFTQKS-SEFEFR---PPASNMYVYAEGLGK--IRSE 139
QY 169 QPWSFQ-----EATQDNFSSGK-----MMKTSSNSQSFSEIASVOTNHS 212
DB 140 PVHFQGGHSGHSPSSISDAAGSSSELSRPTPCQMTPTSDIPAGSQEESIQTSQN 199
QY 213 NGFQSDYGNYPPOSQTLRSRSDGYNWRKYGQKQVKGSENPFSYKTYPCNCTKKKVER 272
DB 200 DS-----RGSTPSILADGYNWRKYGQKVGSENPFSYKTYPCNCTKKKVER 249
QY 273 SLGQITEIVYKTHNHPKQNTNRNSNSS-----LAIPHNSIRTEIPDQSYATHGSG 328
DB 250 SHDQITDIYKTHDHPKQPGRRNSGCGMAAQBRLDKYPSSTGRDEKSGVY--NLN 307
QY 329 QMSAATPNSISIGDDPFSQSKCKSGGDEDEPDAKWIKGENEGHSAPGSR 388
DB 308 PNEGTPNPEVPPISASDGGGAAAS--NRNKDEPDDDDPFKRRMEGAM--ITPLVKP 363
QY 389 VREPRVVVQTTSDIILDDGYRWRKYGQKVGKGNPNRYSYKCTHPCPVKRVKHERASHD 448
DB 364 IREPRVVVQTLSEVDIILDDGYRWRKYGQKVGKGNPNRYSYKCTHACPVKRVKHERASHD 423
QY 449 LRAVITYEGHNHVPARGSGSHSVN---RP-----MPNNAASHN 488
DB 424 PKAVITYEGKHDHVPTRSKSSNHEIQPRFRPDDETDTISLNLGVGISDGPNHASNE-- 481
QY 489 TAATSVRLLPVTHQSDNLSQNRQAPGEGSPFTLEMLQSPGSGFGNPMQSYVQ 548
DB 482 -----HQQN--QQLVNTQHPNGV-----NFRFV-HASPMSSYYAS 514
QY 549 QQLSDNVFSR-TKEPRD 566
DB 515 LNSGMNQYQRETQNETQN 533

RESULT 4
ID_WR26_ARATH STANDARD; PRT; 309 AA.
AC Q9C5T3; Q8GYK8; Q9LYQ5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 26 (WRKY DNA-binding protein 26)
DE (SPF1-like protein)
DE WRKY26 OR AT5G07100 OR MQJ9.27 OR T28J14_40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;

RN SEQUENCE FROM N.A., AND INDUCTION.
RX MEDLINE=21342502; PubMed=11449049;
RA Yu D., Chen C., Chen Z.;
RT "Evidence for an important role of WRKY DNA binding proteins in the
RL regulation of NPR1 gene expression.";
RN Plant Cell 13:1527-1540(2001).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=98344145; PubMed=9679202;
RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
RT features of the regions of 1,381,565 bp covered by twenty one
RT physically assigned PI and TAC clones.";
RL DNA Res. 5:131-145(1998).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohata M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,
RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
RA Farnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grynoprez B., Zimmermann W.,
RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staveren M., Dirks W., Mooljman P., Klein Lankhorst R.,
RA Weitzenger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarreal R., Gielen J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:823-826 (2000).
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=98344145; PubMed=9679202;
RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayaishizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Transcription factor. Interacts specifically with the W
CC responsive cis-acting element (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- INDUCTION: By salicylic acid.
CC -1- SIMILARITY: Belongs to the WRKY class I family.
CC -1- SIMILARITY: Contains 2 WRKY domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; AF224699; AAK28309.1; -
 CC EMBL; AB010697; BAB11168.1; -
 CC EMBL; AL13652; CAB87266.1; -
 CC EMBL; AY084692; AAM61254.1; -
 CC EMBL; AK117545; BAC42206.1; -
 CC EMBL; T48481; T48481; -
 CC InterPro; IPR003657; WRKY.
 CC Pfam; PF03106; WRKY.2.
 CC PROSITE; PS08111; WRKY.2.
 CC Transcription regulation; DNA-binding; Nuclear protein; Repeat.
 CC FT DNA_BIND 111 176 WRKY 1.
 CC FT DNA_BIND 228 293 WRKY 2.
 CC FT CONFLICT 148 148 L -> P (IN REF. 5).
 CC FT CONFLICT 165 165 I -> F (IN REF. 4).
 CC FT CONFLICT 183 183 S -> P (IN REF. 1).
 CC SEQUENCE 309 AA; 34910 MW; B83AF11B93F3909E CRC64;

Query Match 29.08; Score 884; DB 1; Length 309;
 Best Local Similarity 47.74; Pred. No. 2.6e-42;
 Matches 198; Conservative 30; Mismatches 55; Indels 132; Gaps 13;
 QY 58 VPKFKSTPPSLPSPSPSSYFAIPGLSPAEILDSPVLLNSSLPSPTTGAFVAQ 117
 DB 10 VPKFKTATP-----SPLPSPSPYFTWPGULTPADFLDSPLLFTSSNLSPTTGTPPAQ 64
 QY 118 SPNWKSSGG---NQIVKBEKDSFNSFPQTRSGPPASSTATYQSSNVTVTQQPMSFQ 174
 DB 65 SLNY--NNGLLIDKNEIKYEDTT-----PP-----LFLPSMVT---QPLPQL 102
 QY 175 EATKQDNFSGKGMKMTENSSMQSPSPETASVOTNHSNGFSQDYGNYPPQSOTLERRSD 234
 DB 103 DLFKSEIMSNK-----
 QY 235 DGVNWKYGGQVKGNSPRSYKYCTYPCNPTKKKVERSL-DGQITVIVYKTHNPKPO 293
 DB 118 DGVNWKYGGQVKGNSPRSYKYCTYPCNPTKKKVERSLVKGQMIIVYKGNHHPKO 177
 QY 294 NTRNNSNSSLAIPIHNSINTEIPDOSYATHGSGQMSAATPENSISISIDDDFEQSSQ 353
 DB 178 STKSSSTAIA---AHQNS-----SNGD----- 197
 QY 354 KCKSGGDEYDEPDARWKIEGNEGMSAPGSRVTREPRVVQTTSDIILDDGYRWK 413
 DB 198 -----GKDIGDETEAKRWKE-EN-----VKEPRVVQTTSDIILDDGYRWK 241
 QY 414 YGQKVKGNNPRSYKYCTHGPCVPRKVERASHDLRAVITVYEGKHNDVPAAR 468
 DB 242 YGQKVKGNNPRSYKYCTTGCGFVKHVERAFODPKSVITVYEGKHGHQIPTPR 296

RESULT 5

WRK4_ARATH STANDARD; PRT; 514 AA.
 ID AC QX190; Q93WN8; Q9LWGI;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable WRKY transcription factor 4 (WRKY DNA-binding protein 4).
 GN WRKY4 OR ATG13960 OR F7A19.5 OR F16A14.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A., AND INDUCTION.
 RX MEDLINE=21342502; PubMed=11449049;
 RA Yu D., Chen C., Chen Z.;

RT "Evidence for an important role of WRKY DNA binding proteins in the
 RT regulation of NPR1 gene expression.";
 RL Plant Cell 13:1527-1540 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Flower;
 RA Ulker B., Kushnir S., Somssich I.E.;
 RT "Arabidopsis thaliana transcription factor WRKY4.";
 RN Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:816-820 (2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGEC).";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP TISSUE SPECIFICITY, AND INDUCTION.
 RX MEDLINE=21580393; PubMed=1172756;
 RA Robatzek S., Somssich I.E.;
 RT "A new member of the Arabidopsis WRKY transcription factor family,
 RT AtWRKY6, is associated with both senescence- and defence-related
 RT processes.";
 RL Plant J. 28:123-133 (2001).

CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
 CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
 CC responsive cis-acting element (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: In young, mature and senescent leaves.
 CC -!- INDUCTION: By salicylic acid and strongly during leaf senescence.
 CC -!- SIMILARITY: Contains 2 WRKY domains.
 CC -!- CAUTION: Ref.3 sequences differ from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF224703; AAK28313.1; -
 CC EMBL; AF425835; AAL13048.1; -
 CC EMBL; AC007576; AAD39282.1; ALT_SEQ.
 CC EMBL; AC068197; AAF79402.1; ALT_SEQ.
 CC EMBL; AY045676; AAK74034.1; -
 CC EMBL; BT002629; AAO11545.1; -
 CC PIR; G86272; G86272.
 CC InterPro; IPR003657; WRKY.

DR Pfam; PF031106; WRKY, 2.
 DR PROSITE; PS08111; WRKY, 2.
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
 FT DNA_BIND 223 287 WRKY 1.
 FT DNA_BIND 403 468 WRKY 2.
 FT CONFLICT 112 147 AVLDLICI (IN REF. 1).
 FT SEQUENCE 514 AA; 55815 MW; 01010F8745C420C5 CRC64;
 Query Match 28.2%; Score 861.5; DB 1; Length 514;
 Best Local Similarity 38.3%; Pred. No. 8.1e-41;
 Matches 217; Conservative 62; Mismatches 139; Indels 149; Gaps 21;
 QY 31 SPSDLLASPLDNNKPPQGLS-----ERTGSGV-----PKEKSTPPSLPLSPPP 75
 DB 60 SFSOLLGAMSSPTATAAAAAATASYDRGEGTSSSGDVPDRFKQNRPTGLMISQSQ 119
 QY 76 ISPSYFAIPPLGSPAEILLDSFVLLNSNLTSPSTTGAFVAQSPFNKSSGCGNQIVKEE 135
 DB 120 -SP-SMTVPPLGSPAMLLDSPLG-----LSPVQGSY-----GWTQQAQAQ- 162
 QY 136 DKSPNSFQTRSGPPASSTATYQSSNVVQTQPSFQBATKQDNFSSGKGMKNTSS 195
 DB 163 -----VTAQAVQANANNQPTQTEP-----PPSQVSPSSGQAQIPTSPAL 202
 QY 196 SMQSFSPFIASVQTHNSNGFSQSDYGNYPQSQOTLSRSDGYNWRKYGQKOVKGSENP 255
 DB 203 PAQRETSQVTLIE-----HRSQPLNVDPKADDDGYNWRKYGQKOVKGSEFP 250
 QY 256 YKCTYPCPTKKKVERSLDQITEIVYKGFHNHPKQNTRR-NSSNSSSLAIPHNSIR 314
 DB 251 YKCTNPGCPVKKVERSLDQVTEIYKGFHNHPKQNTKRGKNDTANI---NGSSIN 307
 QY 315 TEIPDQSVATHGQMSQATPENSISIGDDPFEQSSQK-----KSGDDEY----- 362
 DB 308 N-----NRGSELGASQPTNSNKTREQHEAVSQATTEHLSEASDGEVNGET 359
 QY 363 -----DEDEPDAKWKLEGE-NEGMSAPGSRVTREPVVQTTSDIILDDGVWRKYG 416
 DB 360 DVREKDNPPKRRSTRSEVRISEPAASHRTVTPEIIVTTSEVLLDDGVWRKYG 419
 QY 417 KVKGNPNRPSYKCTHPGCPVRKHVERASHDLRAVITTEGKINHDPVPAARGSGSHVN 476
 DB 420 KVKGNPNRPSYKCTHPGCPVRKHVERAATDPKAVTTVEGKINHDLPAK-SSSH--- 475
 QY 477 RMPNPNASHTNTAATSVRLLPVTHQSDNSLQNSQAPPEGQSPFTLEMLQSPGSPFS 536
 DB 476 -----AAAAAQLRP-----DN-----RPG----- 489
 QY 537 GFGNPMQSYNQQLSDNVSSRTKEE 563
 DB 490 GLAN-----LNQOQOQOPVARLRUKEE 511
 RESULT 6
 ID WRKY_ARATH STANDARD; PRT; 687 AA.
 AC Q9FG77;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable WRKY transcription factor 2 (WRKY DNA-binding protein 2).
 GN WRKY2 OR AT5G56270 OR MKK23.1 OR K24C1.9.
 OS Arabidopsis thaliana (Mousse-eare crees).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID:3702;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=cv. Columbia;
 TISSUE=Flower;
 Ulker B., Kuehnir S., Somasich I.E.;
 "Arabidopsis thaliana transcription factor WRKY2.";

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC MEDLINE=21016721; PubMed=11130714;
 RX Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsuura A., Muraki A., Nakayama S.,
 RA Nakazaki N., Naruo K., Okumura S., Shingo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
 RA Haberman K., Murray J., Johnson D., Rohlfing T., Nelson J.,
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
 RA Balter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Stromatt C.,
 RA Wagner-McPherson C., Wollam A., Yeakam M., Bell M., Dedhia N.,
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
 RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
 RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S., Zimmermann W.,
 RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Peters S.,
 RA Ramsperger U., Wedler H., Balke K., Wedler E., Klein Lankhorst R.,
 RA van Staveren M., Dirkee W., Mooijman P., Berneiser S., Hempel S.,
 RA Weitzsaecker T., Bothe G., Rose M., Hauf J., Gienel J., Ardiles W.,
 RA Feldpausch M., Lanberth S., Villarroel R., Gienel J., Ardiles W.,
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:823-826 (2000).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RA "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGEC).";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 [4]
 RN TISSUE SPECIFICITY.
 RP MEDLINE=21580393; PubMed=11722756;
 RX Robatzek S., Somasich I.E.;
 RA "A new member of the Arabidopsis WRKY transcription factor family,
 RT AtWRKY6, is associated with both senescence- and defence-related
 RT processes.";
 RL Plant J. 28:123-133 (2001).
 CC -1- FUNCTION: Transcription factor. Interacts specifically with the W
 CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
 CC responsive cis-acting element (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: Low expression in senescent leaves.
 CC -1- SIMILARITY: Belongs to the WRKY class I family.
 CC -1- SIMILARITY: Contains 2 WRKY domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF418308; AL113039.1; -;
 CC EMBL; AB026656; BAB08871.1; -;
 CC EMBL; AY096493; AAM20132.1; -;
 CC EMBL; AY123006; AAM67539.1; -;
 CC InterPro: IPR003657; WRKY.
 CC Pfam; PF03106; WRKY, 2.
 CC PROSITE; PSS0811; WRKY, 2.
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 437 446 POLY-GLY.
 FT DOMAIN 267 331 WRKY 1.

FT DNA_BIND 481 546 WRKY 2.
SQ SEQUENCE 687 AA; 74561 MW; D47EAB1FB0C6335F CRC64;

Query Match 27.9%; Score 852; DB 1; Length 687;
Best Local Similarity 37.1%; Pred. No. 3.8e-40;
Matches 229; Conservative 63; Mismatches 146; Indels 180; Gaps 24;

QY 47 QGGLSE-----RTGSGVVKPKSTPPSLPLSPSPSYFAI-----PPGLSP 90
DB 72 RGLGSRIAARAGFNAPR-----LNTNRTNTDFSIDNLSRPLCTITSSPGLSP 121
QY 91 AELLSPVLLNSNLPSPPTTCFVAQSFNWKSSGGN---QQIVKEE---DKSFNSFQ 145
DB 122 ATLLSPVFLSNPLAQPSPPTG-----KPPFLPGVNGALSKAKDEFFDDIGASFSFH 176
QY 146 TRSGPPASSTAYQSSNVTVTQTPWFSQEQATKDNFSGKGMKMTENSSMQSFSPEIA 205
DB 177 ----PVSRSSEF-----FQGTTEMSVDVGYNNRSSHQSAEBVKGSE 218
QY 206 SVOTNHSNGFQSDYGNYPQSOTL-----GRRSD-----DGY 237
DB 219 NIENSNLYGIETD---NQNGKNTSDVTNTSLETVDHQEBEEQRRGDSMAGGAPABDGY 276
QY 238 NWRKYGOKVKGSENPSPRYKCTYCNCTKKKVERSLEDQITVYKGTNHPKQNTTR 297
DB 277 NWRKYGOKLVKGSENPSPRYKCTNPNQVKKKVERSREGHITIIYKGAHNLKPPPNRR 336
QY 298 N-----SSNSSSLAIPHSNSIRTEIPDQSYAT 324
DB 337 SGMVDVTEFVEQQQQORDSAATWVCNNTQQGGSNENV---EBSGTFEYGNQSGSI 393
QY 325 HG---SQQMDs-----AATPENSISIGDDDFQSQCKKS-----GGDEYDEPDPAKR 371
DB 394 QAQTGGQYSGDPVNVVDASSTFSFSDNEDDRGTHGVSGLVYDGGGGGGGGGDESESKR 453
QY 372 WKIEGENGMSAPGSTRVPRVVVQTTSDIDLDGHRWRYKQVKKVGNPNPRSYKC 431
DB 454 RLEAFAPAEEMSG-STRAIREPRVVQTTSDVDILDGGRWRYKQVKKVGNPNPRSYKC 512
QY 432 THPGCPVRKHVRASHDLRAVITTYEGKHNDVPAAR---GSGSHSVNRPMPNNAENHT 487
DB 513 TAPGCTVRKHVRASHDLKSVITTYEGKHNDVPAARNSSHGGGDSGN-----GNSG 565
QY 488 NTAATSVRLLPVHOSDNLQNRSQAPPEG-----QSPFTLEMLQSPGSF-----533
DB 566 GSAAVS-----HHYNGHSEPPRGFRDQVTTNNQSPFS-----RPFSPQPHL 609
QY 534 ----GFSGFENPMQSVVN 547
DB 610 GPPSGFS-FGLGQTGLVN 626

RESULT 7
WRK3 ARATH STANDARD; PRT; 513 AA.
ID WRK3 ARATH AC Q9ZQ70;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 3 (WRKY DNA-binding protein 3).
GN WRKY3 OR AT2G03340 OR T4M8.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RX MEDLINE=21342502; PubMed=11449049;
RA Yu D., Chen C., Chen Z.;
RT "Evidence for an important role of WRKY DNA binding proteins in the
RT regulation of NPR1 gene expression.";
RL Plant Cell 13:1527-1540(2001).

FT DNA_BIND 481 546 WRKY 2.
SQ SEQUENCE 687 AA; 74561 MW; D47EAB1FB0C6335F CRC64;

Query Match 27.9%; Score 852; DB 1; Length 687;
Best Local Similarity 37.1%; Pred. No. 3.8e-40;
Matches 229; Conservative 63; Mismatches 146; Indels 180; Gaps 24;

QY 47 QGGLSE-----RTGSGVVKPKSTPPSLPLSPSPSYFAI-----PPGLSP 90
DB 72 RGLGSRIAARAGFNAPR-----LNTNRTNTDFSIDNLSRPLCTITSSPGLSP 121
QY 91 AELLSPVLLNSNLPSPPTTCFVAQSFNWKSSGGN---QQIVKEE---DKSFNSFQ 145
DB 122 ATLLSPVFLSNPLAQPSPPTG-----KPPFLPGVNGALSKAKDEFFDDIGASFSFH 176
QY 146 TRSGPPASSTAYQSSNVTVTQTPWFSQEQATKDNFSGKGMKMTENSSMQSFSPEIA 205
DB 177 ----PVSRSSEF-----FQGTTEMSVDVGYNNRSSHQSAEBVKGSE 218
QY 206 SVOTNHSNGFQSDYGNYPQSOTL-----GRRSD-----DGY 237
DB 219 NIENSNLYGIETD---NQNGKNTSDVTNTSLETVDHQEBEEQRRGDSMAGGAPABDGY 276
QY 238 NWRKYGOKVKGSENPSPRYKCTYCNCTKKKVERSLEDQITVYKGTNHPKQNTTR 297
DB 277 NWRKYGOKLVKGSENPSPRYKCTNPNQVKKKVERSREGHITIIYKGAHNLKPPPNRR 336
QY 298 N-----SSNSSSLAIPHSNSIRTEIPDQSYAT 324
DB 337 SGMVDVTEFVEQQQQORDSAATWVCNNTQQGGSNENV---EBSGTFEYGNQSGSI 393
QY 325 HG---SQQMDs-----AATPENSISIGDDDFQSQCKKS-----GGDEYDEPDPAKR 371
DB 394 QAQTGGQYSGDPVNVVDASSTFSFSDNEDDRGTHGVSGLVYDGGGGGGGGGDESESKR 453
QY 372 WKIEGENGMSAPGSTRVPRVVVQTTSDIDLDGHRWRYKQVKKVGNPNPRSYKC 431
DB 454 RLEAFAPAEEMSG-STRAIREPRVVQTTSDVDILDGGRWRYKQVKKVGNPNPRSYKC 512
QY 432 THPGCPVRKHVRASHDLRAVITTYEGKHNDVPAAR---GSGSHSVNRPMPNNAENHT 487
DB 513 TAPGCTVRKHVRASHDLKSVITTYEGKHNDVPAARNSSHGGGDSGN-----GNSG 565
QY 488 NTAATSVRLLPVHOSDNLQNRSQAPPEG-----QSPFTLEMLQSPGSF-----533
DB 566 GSAAVS-----HHYNGHSEPPRGFRDQVTTNNQSPFS-----RPFSPQPHL 609
QY 534 ----GFSGFENPMQSVVN 547
DB 610 GPPSGFS-FGLGQTGLVN 626

RESULT 7
WRK3 ARATH STANDARD; PRT; 513 AA.
ID WRK3 ARATH AC Q9ZQ70;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 3 (WRKY DNA-binding protein 3).
GN WRKY3 OR AT2G03340 OR T4M8.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RX MEDLINE=21342502; PubMed=11449049;
RA Yu D., Chen C., Chen Z.;
RT "Evidence for an important role of WRKY DNA binding proteins in the
RT regulation of NPR1 gene expression.";
RL Plant Cell 13:1527-1540(2001).

FT DNA_BIND 481 546 WRKY 2.
SQ SEQUENCE 687 AA; 74561 MW; D47EAB1FB0C6335F CRC64;

Query Match 27.9%; Score 852; DB 1; Length 687;
Best Local Similarity 37.1%; Pred. No. 3.8e-40;
Matches 229; Conservative 63; Mismatches 146; Indels 180; Gaps 24;

QY 47 QGGLSE-----RTGSGVVKPKSTPPSLPLSPSPSYFAI-----PPGLSP 90
DB 72 RGLGSRIAARAGFNAPR-----LNTNRTNTDFSIDNLSRPLCTITSSPGLSP 121
QY 91 AELLSPVLLNSNLPSPPTTCFVAQSFNWKSSGGN---QQIVKEE---DKSFNSFQ 145
DB 122 ATLLSPVFLSNPLAQPSPPTG-----KPPFLPGVNGALSKAKDEFFDDIGASFSFH 176
QY 146 TRSGPPASSTAYQSSNVTVTQTPWFSQEQATKDNFSGKGMKMTENSSMQSFSPEIA 205
DB 177 ----PVSRSSEF-----FQGTTEMSVDVGYNNRSSHQSAEBVKGSE 218
QY 206 SVOTNHSNGFQSDYGNYPQSOTL-----GRRSD-----DGY 237
DB 219 NIENSNLYGIETD---NQNGKNTSDVTNTSLETVDHQEBEEQRRGDSMAGGAPABDGY 276
QY 238 NWRKYGOKVKGSENPSPRYKCTYCNCTKKKVERSLEDQITVYKGTNHPKQNTTR 297
DB 277 NWRKYGOKLVKGSENPSPRYKCTNPNQVKKKVERSREGHITIIYKGAHNLKPPPNRR 336
QY 298 N-----SSNSSSLAIPHSNSIRTEIPDQSYAT 324
DB 337 SGMVDVTEFVEQQQQORDSAATWVCNNTQQGGSNENV---EBSGTFEYGNQSGSI 393
QY 325 HG---SQQMDs-----AATPENSISIGDDDFQSQCKKS-----GGDEYDEPDPAKR 371
DB 394 QAQTGGQYSGDPVNVVDASSTFSFSDNEDDRGTHGVSGLVYDGGGGGGGGGDESESKR 453
QY 372 WKIEGENGMSAPGSTRVPRVVVQTTSDIDLDGHRWRYKQVKKVGNPNPRSYKC 431
DB 454 RLEAFAPAEEMSG-STRAIREPRVVQTTSDVDILDGGRWRYKQVKKVGNPNPRSYKC 512
QY 432 THPGCPVRKHVRASHDLRAVITTYEGKHNDVPAAR---GSGSHSVNRPMPNNAENHT 487
DB 513 TAPGCTVRKHVRASHDLKSVITTYEGKHNDVPAARNSSHGGGDSGN-----GNSG 565
QY 488 NTAATSVRLLPVHOSDNLQNRSQAPPEG-----QSPFTLEMLQSPGSF-----533
DB 566 GSAAVS-----HHYNGHSEPPRGFRDQVTTNNQSPFS-----RPFSPQPHL 609
QY 534 ----GFSGFENPMQSVVN 547
DB 610 GPPSGFS-FGLGQTGLVN 626

RESULT 7
WRK3 ARATH STANDARD; PRT; 513 AA.
ID WRK3 ARATH AC Q9ZQ70;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 3 (WRKY DNA-binding protein 3).
GN WRKY3 OR AT2G03340 OR T4M8.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RX MEDLINE=21342502; PubMed=11449049;
RA Yu D., Chen C., Chen Z.;
RT "Evidence for an important role of WRKY DNA binding proteins in the
RT regulation of NPR1 gene expression.";
RL Plant Cell 13:1527-1540(2001).

FT DNA_BIND 481 546

```

238 HRSQPNQADPADDGYNWRKYGQKVGSGDFPRSYKCTHPACPVKKKVERSLLDQGVTEI 297
282 VYKTHNHPKQNTNRSSNSSSLAIPHSNSIRTEIPDQSYATHGSGQMSAATPENSSI 341
298 IYKQHNHLPKQNGNNNGCKS-----SDIANQ-FQTSNSLNKSKRDQETSQV 346
342 SIGDDDFEQSSQK-----CKSGGDEYDEPDKAEKWKTEGENEGMSAPGSRVREPRVVQ 397
347 TTTEQMSAQSSEVGNNAETSVGERHEDEPDKRNTEVRSEVPVASSHTVTEPRILVQ 406
398 TTSDIDILDCGYRWRKYGQKVGKGNPNRPSYKCTHPGCPVRKHVERASHDLRAVITYE 457
407 TTSEVDLLDGYRWRKYGQKVGKGNPYRPSYKCTTDPDCGVRKHVERAATDPRAVITYE 466
458 GKNHVDVPAARGSGSHSVNRPMPNNAHNNT 489
467 GKNHVDVPAARTS-SHQLR-----PNNQHN-TST 493

RESULT 8
WR34 ARATH
ID WR34 ARATH STANDARD; PRT; 568 AA.
AC O65590;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 34 (WRKY DNA-binding protein 34).
GN WRKY34 OR AT4G26440 OR M3E9.130.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delsen Y., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hehese J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Billam L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirckx W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneser S., Hempel S., Feldpausch M., Lambert S., van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzi A.,
RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,

```

```

RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stonking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Pante M., Pepin K., Hillier L.,
RA Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Beighoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Grant S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999);
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
CC responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the WRKY class I family.
CC -!- SIMILARITY: Contains 2 WRKY domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.iesb-sib.ch/announce/
CC or send an email to license@iesb-sib.ch).
CC -----
CC EMBL; AY052649; AAL11010.1; -
CC EMBL; AL022223; CAA18226.1; -
CC EMBL; AL161565; CAB79499.1; -
CC PIR; T05060; T05060.
CC InterPro; IPR003657; WRKY.
CC Pfam; PF03106; WRKY; 2.
CC PROSITE; PS50811; WRKY; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DNA_BIND 172 236 WRKY 1.
FT DNA_BIND 366 431 WRKY 2.
SQ SEQUENCE 568 AA; 62166 MW; 31F350AA46341F7B CRC64;
Query Match 24.3%; Score 742; DB 1; Length 568;
Best Local Similarity 37.3%; Pred. No. 3.6e-34;
Matches 193; Conservative 55; Mismatches 119; Indels 150; Gaps 16;
QY 86 PGLSPAEILDSPVLLNSSNLPSPPTTGAFAVQSFNWKSSGGNQIQIVKEEDKSFNSPQ 145
DB 88 PGLSPATLLESVPVL--SNPLLSPTTGLK-----SSVPSDKAKAELFDDITSLAQ 137
QY 146 TRGSPASSTATTQSSNVTVTQTPNSPQATKQDNFSSGKGMKNTENSSMSQSPSEIA 205
DB 138 TISGSGLDPT-----NIALEPDDSDQYBERQ-----LGLGLG-----DSMACCAP--- 176
QY 206 SVQTNHNSGQSDYGVNVPQSTLSRRSDGYNWRKYGQKVGSGKGNKTPNSRYKCTPNCP 265
DB 177 -----ADGYNWRKYGQKLVKSGSEYPSRYKCTPNCE 209
QY 266 TKKKVERSLDQITEIVYKGTHTNHPKQNTNRSSNSSSLA----- 306
DB 210 AKKKVERSRBGHIIETITGDHTSKPPPPNRRRGISGSGTGQDMQIDATEYEGFAGTNN 269
QY 307 IPHSNSIRTEIPQSYA-----THGSGQMSAATPENSSISIGDDDFEQSSQCKSG 358
DB 270 IEWTSFVSAELEYGSHSGSMQVNGTHQFGYGAAA-----DALYRDENEDRTSHMSVSL 325
QY 359 GDEYDEDEPDKAEKWKTEG-----ENEGMSAPGSRVREPRVVVQTTSDIDILDDGYRWRKY 415
DB 326 TYGGEVESESKRKLEAVATETSG-----STRASRPRVVVQTTSDIDILDDGYRWRKY 381
QY 416 QKVVGKGNPNRPSYKCTHPGCPVRKHVERASHDLRAVITYEGKHNHVDVPAAR-----GS 470

```



```

||||| 382 QKVGKGNPRSYKCTANGCTTGVKVERASDDFKSLVLTYYIGKTHVVPAARNSSHVGA 441
||||| 471 GS-----HSVNRPMNPNASHTNTAATSVRLLPVTHQSDNSLQNRQOAPP 516
||||| 442 GSGGTGQSLATQTHNHNHYPHRSRSEGLATA-----NSSLDFDQSH--- 485
||||| 517 EGQSPFTLEMLQSPGFGFGNPMQSYVNOQQLSD 553
||||| 486 -----LRHTGFSV-----YIGQSELSD 503

RESULT 9
ID WR58 ARATH STANDARD; PRT; 423 AA.
AC Q93WU7; O9MAC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Probable WRKY transcription factor 58 (WRKY DNA-binding protein 58).
GN WRKY58 OR AT3G01080 OR T4P13.24.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RA Kushnir S., Ulker B., Somsich I.E.;
RT "Arabidopsis thaliana transcription factor WRKY58.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salancoubat M., Lemcke K., Rieger M., Ansoerge W., Unseled M.,
RA Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choine N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Catolico L., Weissbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Floss H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Schaefer M., Schoen O., Bargues M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
RA Mannheim G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.-L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato I., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana.";
RL Nature 408:820-822(2000).
CC -!- FUNCTION: transcription factor. Interacts specifically with the w
box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Contains 2 WRKY domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; AF426254; AAL29431.1; -.
DR EMBL; AC008261; AAF26166.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DNA BIND 161 225 WRKY 1.
FT DNA BIND 300 365 WRKY 2.
FT DOMAIN 150 154 POLY-ASN.
FT DOMAIN 257 270 POLY-ASP.
FT CONFLICT 81 81 L->LNGTMSVSPGGGRSTAGMPAGGPMFTIPSGFSPS
FT FT
SQ SEQUENCE 423 AA; 47121 MW; EF6C2F1BF3F16B15 CRC64;
Query Match 24.2%; Score 738; DB 1; Length 423;
Best Local Similarity 37.8%; Pred. No. 4.2e-34;
Matches 182; Conservative 52; Mismatches 143; Indels 104; Gaps 14;
QY 65 PPSPLPLSPPISSSYF-----AIPPLGPAELLDSPVLLNSNLLPSPTTGAFVAQ 117
DB 21 PTTTVPHRPFAJETAAYFFGGGGLSLSPG---PLSFVSLFVDNFPDVL----- 67
QY 118 SFNWKSSSGNQOIVKEEDKSPNSFSFQTRSGPPASSTAT-----YQSSNVTVTOQPWS 172
DB 68 -----TPDNQRTT-----SFTQLTSPMFFPQSSAHTGFIQPRQSQ--PQQRPDPT 113
QY 173 FQEAATQDNFSGKGMKMTENSSMSQSFPEIASVQTNHSGFQSDYGNYPQSQRTLSRR 232
DB 114 FPHMPPSTSVAVHGQSLDVS-----QVDQARAHNNPNNNNRNSVNVVVDKP 165
QY 233 SDGYNWRKYGQKQKXGSENPRSYKCTYKCTPKKKKVERSLLDQITETVYKGTNHPKP 292
DB 166 ADDGYNWRKYGQKPIKGCYPRSYKCTHVNCVFKKKKVERSDDQITQIYKQHDHERP 225
QY 293 QNTRNSNSSSLAIPHSNISRTETPDQSYATHGSCQMSAATPENSISIGDDDEQSS 352
DB 226 QNRKGGGGRDS-----TEV-----GGAGQMMESSDGSYRKDHDDDDDD--- 264
QY 353 QKCKSGGDRVDEPDAKRWKKEGENEGMSAPGSRVREPRVVVQTTSIDILDDGVWR 412
DB 265 -----DEDDDELPASKRRIDG-----VSTHRTVTEFKIIVQTKSEVLLDDGTRWR 312
QY 413 KYGQKVVKNPNPRSYKCTHPCGCVRKVERASHDLRAVITYEGKHNHDVPAARGSGS 472
DB 313 KYGQKVVKNPNPRSYKCTTPNCTVRKVERASTDAKAVITYEGKHNHDVPAARG-- 370
QY 473 HSNVRPMNPNASHTNTAATSVRLLPVTHQSDNSLQNRQOAPP 516
DB 371 -----TAAATAAAVGPDSHHRMSGNNNQHMFGNNNNNTQSPVLLR 415
QY 526 M 526
DB 416 L 416
RESULT 10
ID WR44 ARATH STANDARD; PRT; 429 AA.
AC Q9ZU00;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE WRKY transcription factor 44 (WRKY DNA-binding protein 44)
DE (TRANSPARENT TESTA GLABRA 2).
GN WRKY44 OR TTG2 OR AT2G37260 OR P3G5.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```


RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana";
RL Nature 402:761-768 (1999).
CC -1- FUNCTION: Transcription factor. Binds to a 5'-CGTGGACGAG-3'
CC consensus core sequence which contains a w box, a frequently
CC occurring elicitor-responsive cis-acting element.
CC -1- COFACTOR: Requires metal-ions (probably zinc) for its binding
CC activity.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9SI37-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9SI37-2; Sequence=VSP_007124;
CC -1- TISSUE SPECIFICITY: Expressed to similar levels in root and
CC flower, to a somewhat lower level in stem and to low levels in
CC leaf and siliques.
CC -1- MISCELLANEOUS: Binding to target DNA is mediated mainly by the C-
CC terminal WRKY domain, while part of the activation domain is
CC located between positions 210 and 285.
CC -1- SIMILARITY: Belongs to the WRKY class I family.
CC -1- SIMILARITY: Contains 2 WRKY domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X92976; CAA63554.1; -;
DR EMBL; AF442389; AAL35282.1; -;
DR EMBL; AF442390; AAL35283.1; -;
DR EMBL; AC007211; AAD25579.1; -;
DR EMBL; AC006955; AAM15341.1; -;
DR PIR; F84462; F84462.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS0811; WRKY; 2.
KW Transcription regulation; DNA-binding; Zinc; Nuclear protein; Repeat;
KW Alternative splicing; Activator.
FT DNA_BIND 105 169 WRKY 1.
FT DNA_BIND 301 366 WRKY 2.
FT VARSPPLIC 191 214 Missing (in isoform 2).
FT /FTID=VSP_007124.
SQ SEQUENCE 487 AA; 54010 MW; 61721DB016897C38 CRC64;

Query Match 18.2%; Score 556.5; DB 1; Length 487;
Best Local Similarity 35.0%; Pred. No. 5.1e-24;
Matches 131; Conservative 50; Mismatches 98; Indels 95; Gaps 10;

QY 234 DDGNNWRKYGKQVKGSGENRPSRYKCTYPNCPKTKKXVSLDQIETIVYKGTNHPKP- 292
DB 111 EDGNNWRKYGKLVKGNFEVRSYRCTHPNCKAKKQLERSAGGVVDTVYFGSHDPKPL 170
QY 293 -----QNTRRNSNSS-----LAI-----PHSNSIRTEIPDQS 321
DB 171 AGAVPINQDKRSVDFTAVSGEQRIDIVSLYKLCIVSYDIFVEKTSQSVQTLQATEP 230
QY 322 YATHGQMDASATPENSISIGDDDFEQSQKCKGSGDYDEB--EPDAKRWKIEGENE 379
DB 231 PKIHGGLHV-----SVIPPADVDKTDISQSSRITGDNTHKDYNSPTAKRKKGGNIE 282
QY 380 GMSAPGSRVREPRVVTQSDILDGGRWKYGGKVVKGNNPRSPRYKCTHPGCPVR 439
DB 283 --LSPVSTNDGRIVVHTQTLFVNDVGRWKYGGKVVKGNNPRSPRYKCTHPGCPVK 340
QY 440 KHVERASHDLRAVITTYEGKHNDVPAARGSGSHSVNRPMPNPNASHTNTAATSVRLLPV 499

DB 341 KHVERSSHTKLLITTYEGKHNDMP-----PGRVVTHNN----- 375
QY 500 IHQSDNLSLONRSQAPPEGQSPTLEMLQSPGSGFGNGFNGPMQSVVQQLSDNVFSSR 559
DB 376 --MLDSEVDDKEGDANKTFQS-----STLQSTKQHQHVEDHL----- 410
QY 560 TKREPRDDMFLESL 573
DB 411 -RKKTGTNGFEKSL 423

RESULT 12
WR19_ARATH
ID WR19_ARATH STANDARD; PRT; 1895 AA.
AC Q9SZ67;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 19 (WRKY DNA-binding protein 19).
GN WRKY19 OR AT4G12020 OR F16J13.90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
ON NCBI_TaxID=3702;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RC MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoerge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
RA Kreis M., Delsen Y.M., Putgdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzengger T., Bothe G., Ramepger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkee W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McLeay K., Mayes R.,
RA Pettett A., Rajadream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Partmann B., Grandeth K., Dauner D., Herzl A.,
RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bedian M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;

"Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana";
 Nature 402:769-777(1999).
 CC -!- FUNCTION: Transcription factor. Interacts specifically with the W box (5'-(T)TCAC(C/T)-3'), a frequently occurring elicitor-responsive cis-acting element. May act also as a disease resistance protein with a serine/threonine-protein kinase activity (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Belongs to the WRKY class I family.
 CC -!- SIMILARITY: Belongs to the disease resistance X-TIR-NB-LRR-X family.
 CC -!- SIMILARITY: Contains 7 leucine-rich (LRR) repeats.
 CC -!- SIMILARITY: Contains 1 NB-ARC domain.
 CC -!- SIMILARITY: Contains 1 protein kinase domain.
 CC -!- SIMILARITY: Contains 2 WRKY domains.
 CC -!- DATABASE: NAME-NB-LRRs;
 CC NOTE=Functional and comparative genomics of disease resistance gene homologs;
 CC WWW="http://niblr.ucdavis.edu".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL049638; CAB40943.1; -;
 CC EMBL; AL161533; CAB78245.1; -;
 CC PIR; T06609; T06609.
 CC HSPF; P02876; 9MGA.
 CC InterPro; IPR003593; AAA_ATPase.
 CC InterPro; IPR000767; Disease_resist.
 CC InterPro; IPR006210; IEGF.
 CC InterPro; IPR001611; LRR.
 CC InterPro; IPR002182; NB-ARC.
 CC InterPro; IPR003822; PAH.
 CC InterPro; IPR000719; Prot_kinase.
 CC InterPro; IPR002290; Ser_thr_pkinase.
 CC InterPro; IPR003657; WRKY.
 CC Pfam; PF00560; LRR; 4.
 CC Pfam; PF00931; NB-ARC; 1.
 CC Pfam; PF02671; PAH; 1.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF03106; WRKY; 2.
 CC PRINTS; PR00364; DISEASERST.
 CC ProDom; PD000001; Prot_kinase; 1.
 CC SMART; SM00382; AAA; 1.
 CC SMART; SM00181; EGF; 1.
 CC SMART; SM00220; S_TKc; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
 CC PROSITE; PS50811; WRKY; 2.
 CC Plant defense; transferase, Serine/threonine-protein kinase;
 CC Transcription regulation; Nuclear protein; ATP-binding; DNA-binding;
 CC Repeat; Leucine-rich repeat.
 CC FT DNA_BIND 462 526 WRKY 1.
 CC FT DNA_BIND 635 700 WRKY 2.
 CC FT DOMAIN 800 1087 NB-ARC.
 CC FT REPEAT 1204 1226 LRR 1.
 CC FT REPEAT 1257 1281 LRR 2.
 CC FT REPEAT 1304 1327 LRR 3.
 CC FT REPEAT 1328 1348 LRR 4.
 CC FT REPEAT 1349 1371 LRR 5.
 CC FT REPEAT 1372 1395 LRR 6.
 CC FT REPEAT 1419 1443 LRR 7.
 CC FT DOMAIN 1626 1877 PROTEIN_KINASE.
 CC FT NP_BIND 844 851 ATP (POTENTIAL).
 CC FT NP_BIND 1632 1640 ATP (BY SIMILARITY).
 CC FT BINDING 1654 1654 ATP (BY SIMILARITY).

FT ACT SITE 1758 1758 BY SIMILARITY.
 FT DOMAIN 97 307 GLY-RICH.
 FT DOMAIN 35 39 POLY-SER.
 FT DOMAIN 87 90 POLY-SER.
 FT DOMAIN 980 983 POLY-LEU.
 FT DOMAIN 1568 1571 POLY-SER.
 SQ SEQUENCE 1895 AA; 210320 MW; 1C19D3E164C9363 CRC64;
 Query Match 17.0%; Score 520; DB 1; Length 1895;
 Best Local Similarity 32.9%; Pred. No. 2.6e-21;
 Matches 154; Conservative 72; Mismatches 160; Indels 82; Gaps 17;
 QY 157 TYQ-----SSNVTVQTOQPMFQBEATKQDNFSGKGMKMTENSSSMOSFPEIASVQT 209
 Db 397 TYQQAQTVQANANMQPQTEVP---SSAVQSFSSGQPOIPTS-----APDSLLAK 444
 QY 210 NNSNGFQSDYGVPPQSTLRSDDGYWRKYQKQVKGSGENSPRSYKCTVPCPTKKK 269
 Db 445 SNTSGI-TTIEHMSQQPLNVKQVNDGYNMQYQKVKVGSFPLSYKCTYLGCPSKRR 503
 QY 270 VERSLDGQITEIVYKGTNNHPKQNTNRSSNSSSLAIPHSNSIRTEIPDQSVATHGS-- 327
 Db 504 VERSLDGQVAEIVYKDRHNEHPNQKOGSTTYLSSGSHINCKMSSELTAQSFSSNKTKI 563
 QY 328 GQMDSAATPEN-SSISIGDDPEQSKCKSGGDEYDEDEPAKRWKIEGENGMSAPGS 386
 Db 564 EQEAAASLATTIEYMSASDNEEDSNGTSEG--EKDEDEPPKRRITETVQVSELADSD 621
 QY 387 RTVREPRVVQTTSDIILDDGYWRKYQKQVKGKGNPNR-----SYKCTHGPCVPRKH 441
 Db 622 RTVREPRVIFQTTSEVDNLDGYWRKYQKQVKGKGNPYPRFSSSKDYYDVIRYGRADISN 681
 QY 442 VERASHDLRAV-----ITTYEGKHND-VPAARGSGSHSVNRPMNPNASNHTNTAATSVR 495
 Db 682 EDFISH-LRASLCRRGISVTEKFEVDALPKRVLIIVTSTVPSNLLNILEHQHTEDR 740
 QY 496 LL-PVIHQSD-----NSLQNR--SQAPPEGQSPFTLEMLQSPGSGFGFNGPMQSYV 546
 Db 741 VVYPIFYRLSPYDFVCNSKNYERFYLDPEKKWQAALKEITQMPG-----DMFLE 575
 QY 547 NQOQLSDNVFSSRTKEPRD-----DMFLE-----SLLC 575
 Db 786 --YTLTKSESELIDEIVRDALKVLCGADKVNIMGMQVVEILSLLC 831
 RESULT 13
 WR32 ARATH STANDARD; PRT; 466 AA.
 AC P59583; O65556;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable WRKY transcription factor 32 (WRKY DNA-binding protein 32).
 GN WRKY32 OR AT4G30935 OR F6118.4.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoerge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidtheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohnsels J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J., Vandenbussche P.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Aert R., Defoor E.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,

RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLaughlin K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loecherer T.-H.,
RA Dose S., de Haan M., Maaree A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenet O., Quigley F., Clabaud G., Muehlen A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Bielek C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sakon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.,
RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RA thaliana.";
RA Nature 402:769-777(1999).
RA [2]
RA SEQUENCE FROM N.A.
RA STRAIN=cv. Columbia; TISSUE=Flower;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RA SSP consortium (Salk/Stanford/PGEC).";
RA Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RA -!- FUNCTION: Transcription factor. Interacts specifically with the W
RA box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
RA responsive cis-acting element (By similarity).
RA -!- SUBCELLULAR LOCATION: Nuclear (Probable).
RA -!- SIMILARITY: Belongs to the WRKY class I family.
RA -!- CAUTION: Ref.1 sequences differ from that shown due to erroneous
RA gene model prediction. AT4G30930 and AT4G30935 were originally
RA fused into a single gene.
RA -----
RA This SWISS-PROT entry is copyright. It is produced through a collaboration
RA between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA the European Bioinformatics Institute. There are no restrictions on its
RA use by non-profit institutions as long as its content is in no way
RA modified and this statement is not removed. Usage by and for commercial
RA entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
RA or send an email to license@isb-sib.ch).
RA -----
RA EMBL; AL022198; CAA18200.1; ALT_SEQ.
RA EMBL; AL161578; CAB79811.1; ALT_SEQ.
RA EMBL; BT004086; AAO42113.1; -.
RA PROSITE; PSS0811; WRKY; 2.
RA Transcription regulation; DNA-binding; Nuclear protein; Repeat.
RA DNA BIND 162 226 WRKY 1.
RA DNA BIND 325 390 WRKY 2.
RA SEQUENCE 466 AA; 51480 MW; A4CC4E13B5262DAF CRC64;
RA -----
RA Query Match 15.8%; Score 482; DB 1; Length 466;
RA Best Local Similarity 39.3%; Pred. No. 6.2e-20;
RA Matches 118; Conservative 39; Mismatches 111; Indels 32; Gaps 10;

QY 204 IASVOTNHSNGFQSDYGN---YPPQSQTLRSRDDGYNWRKYQKQKQVKGSENPRSYKCT 260
DB 138 LVSVPTKQQRSDSPVNRSLSTVTPRTPAR---DGYNRKYQKQKQVKGSGRSYRCT 194
QY 261 YNCPYTKKVERSLD-GQTEIVYKTHNHPKQPNTRNRSNSSLAIAPHSNIRTEIPD 319
DB 195 YTEC-CAKKIECSNDGNGVVEIVNKLHTEHPRTK-----SFSPEIRVTTAIRPVSD 248
QY 320 QSVATH---GSGQMDSAATPENSSISIGDDPEQSQKCKGSGGDEYDEPDKEWKIE 375
DB 249 DTVEELSVPGSDPSASTKEYICES---QTLVDRKRHCENEAVS---EPEPKR-RLK 300
QY 376 GNEGNSAPGSRTRVPRVVVQTTSDIDLDGYNWRKYQKQKQVKGSENPRSYKCTHPG 435
DB 301 KONSQSSDSVSKPGKKNKVFVHAAGDVGICGSGYRWRKYQKQKQVKGSENPRSYRCTSG 360
QY 436 CPVRKIVERASHDLRAVITYYEGKHNVDPAARGSGSHSVNRPMPNNSHTNTAATSVR 495
DB 361 CPVRKHETAVENTKAVIITYYKGVHNDMPVK-----KRGPPSSMLVAAAPTSMR 413

RESULT 14
WR10 ARATH STANDARD; PRT; 506 AA.
ID WR10 ARATH STANDARD; PRT; 506 AA.
AC Q9LGS5; Q8VWQ3;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 10 (WRKY DNA-binding protein 10).
OS WRKY10 OR ATIG55600 OR F20N2.3.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OC NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RC Ulker B., Kushnir S., Sonssich I.E.;
RA "Arabidopsis thaliana transcription factor WRKY10.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RC MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltschker J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Unterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RA thaliana.";
RL Nature 408:816-820(2000).
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
CC responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the WRKY class I family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:52 ; Search time 40 Seconds
(without alignments)

3709.506 Million cell updates/sec

Title: US-09-890-811B-10

Perfect score: 3050

Sequence: 1 MASSSGSLDTSASANSFTNF.....FSSRTKEPRDDMFLESLIC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1822	59.7	549	10 Q40090	Q40090 ipomoea bat
2	1389.5	45.6	559	10 Q941B6	Q941B6 nicotiana t
3	1379.5	45.2	515	10 Q81639	Q81639 pimipinella
4	1342	44.0	514	10 Q9SQ04	Q9SQ04 petroselinu
5	1331	43.6	514	10 Q40827	Q40827 petroselinu
6	1278	41.9	477	10 Q9XJ26	Q9XJ26 nicotiana t
7	1222	40.1	353	10 Q9SXP5	Q9SXP5 nicotiana t
8	1201.5	39.4	378	10 Q941B4	Q941B4 nicotiana t
9	1139.5	37.4	512	10 Q8S8P5	Q8S8P5 arabidopsis
10	1116.5	36.6	501	10 Q9XFF1	Q9XFF1 avena sativ
11	1081	35.4	357	10 Q94D89	Q94D89 oryza sativ
12	1032.5	33.9	402	10 Q38704	Q38704 avena fatua
13	927	30.4	727	10 Q941B3	Q941B3 nicotiana t
14	924	30.3	739	10 Q9ARF7	Q9ARF7 lycopersico
15	922	30.2	739	10 Q9ARD0	Q9ARD0 lycopersico
16	894.5	29.3	387	10 Q94AT4	Q94AT4 arabidopsis

17	894	29.3	309	10 Q8GYK8	Q8GYK8 arabidopsis
18	852	27.9	687	10 Q9FG77	Q9FG77 arabidopsis
19	843	27.6	571	10 Q9LMG1	Q9LMG1 arabidopsis
20	840	27.5	528	10 Q9ZPL6	Q9ZPL6 nicotiana t
21	816.5	26.8	485	10 Q8H1E9	Q8H1E9 arabidopsis
22	812.5	26.6	485	10 Q94AP6	Q94AP6 arabidopsis
23	799	26.2	490	10 Q9SSX9	Q9SSX9 nicotiana t
24	778.5	25.5	509	10 Q39658	Q39658 cucumis sat
25	759.5	24.9	488	10 Q8W1M6	Q8W1M6 retama raet
26	742	24.3	568	10 Q65590	Q65590 arabidopsis
27	680.5	22.3	415	10 Q8SE29	Q8SE29 oryza sativ
28	677	22.2	369	10 Q941B5	Q941B5 nicotiana t
29	613.5	20.1	439	10 Q9AUW5	Q9AUW5 oryza sativ
30	609	20.0	591	10 Q947G7	Q947G7 solanum dul
31	574.5	18.8	463	10 Q43388	Q43388 arabidopsis
32	565.5	18.5	296	10 Q40829	Q40829 petroselinu
33	561.5	18.4	349	10 Q9ZU00	Q9ZU00 arabidopsis
34	556.5	18.2	487	10 Q9SI37	Q9SI37 arabidopsis
35	555	18.2	454	10 Q9ZPL7	Q9ZPL7 nicotiana t
36	548	18.0	482	10 Q8H6W1	Q8H6W1 oryza sativ
37	520	17.0	1895	10 Q9SZ67	Q9SZ67 arabidopsis
38	488.5	16.0	782	10 Q65556	Q65556 arabidopsis
39	390	12.8	513	10 Q9ARF4	Q9ARF4 capsella ru
40	386	12.7	506	10 Q9LG05	Q9LG05 arabidopsis
41	378.5	12.4	485	10 Q8VMQ3	Q8VMQ3 arabidopsis
42	337	11.0	565	10 Q94E13	Q94E13 oryza sativ
43	329	10.8	379	10 Q9FE35	Q9FE35 oryza sativ
44	318	10.4	191	10 Q8S8S5	Q8S8S5 arabidopsis
45	318	10.4	218	10 Q93WY4	Q93WY4 arabidopsis

ALIGNMENTS

RESULT 1

Q40090 PRELIMINARY; PRT; 549 AA.
ID AC Q40090;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE SPF1 protein.
OS Ipomoea batatas (Sweet potato) (Batatae).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=4120;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kokei No.14; TISSUE=Petiote;
RX MEDLINE=95058910; PubMed=7969025;
RA Ishiguro S., Nakamura K.;
RT "Characterization of a cDNA encoding a novel DNA-binding protein,
RT SPF1, that recognizes SP8 sequences in the 5'-upstream regions of
RT genes coding for sporamin and beta-amylase from sweet potato."
RL Mol. Gen. Genet. 244:563-571(1994).
DR EMBL: D30038; BAA06278.1; --
DR TRANSFAC; T03975; --
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS0811; WRKY; 2.
KW DNA-binding.
SQ SEQUENCE 549 AA; 55695 MW; 5EC11DBCFF96FA3A CRC64;

Query Match 59.7%; Score 1822; DB 10; Length 549;
Best Local Similarity 62.0%; Pred. No. 4.4e-120;
Matches 369; Conservative 66; Mismatches 92; Indels 68; Gaps 14;

QY 1 MASSSGSLDTSASANSFTNFSTHPTFTSPFLLASPLDNNKPPQGGI----SRTGS 56

DB 1 MAASSTIDAPTASS---FSFTASSFMSSFTDLIASDAYSGSVSRGDIARTGTS 57

QY 57 GVPKFKSTPPSLPLSPPPISPSSTYFAIPGLSPAEILDSPVLLNSSNLPSTTGAFA 116

```

Db 58 GVFKFKSLPPLSPPLSPSPSPSYFAFPGLSPSELLDSPVLLSSNLLPSPPTGTFPA 117
Qy 117 QSNFWKSSGNGNQIVKEEDKSFNFQFQTRSGPPASSTATYQSSNVTVQTPWFSQEA 176
Db 118 QTNWKNKSNASQDVKEEGYDFDSFQTN-----ASMTLYEDSK----- 160
Qy 177 TKQDNFSGGKMKMTENSSNQSPSPFIASVQTNHSN--GFQSDYGNV--PQSQTL--SR 231
Db 161 -----RKDELNSLQSLPPVTTSTQSSQNNGSGSYEYNNQCCPPSQOTUREQR 207
Qy 232 RSDDGYNWRKYQKQVKGSENPRESYKCTPNCPTKKKVERSILDGQITEIVYKGTNHPK 291
Db 208 RSDDGYNWRKYQKQVKGSENPRESYKCTPNCPTKKKVERALDGQITEIVYKGAHNHPK 267
Qy 292 PONTNRNS-----SNSSLAIPHNSIRTEIPDQSYATHGSGQMDSAATPENSISIGDD 347
Db 268 PQSTRSSSSSTASSASTLAAQSYNAPASDVDPQSYWNGNGQMDSVATPENSISVGDDE 327
Qy 348 FEQSSOKKSGGDEYDEDEPAKWKTEGENEGMSAPGSRVREPRVVVQTTSDIDLDD 407
Db 328 FEQSSQKRESGGDEYDEDEPAKWKTEGENEGMSAPGSRVREPRVVVQTTSDIDLDD 387
Qy 408 GYRWKYGQKVGKGNPNPNRSYKCTHPGCPVRKHVERASHDLRAVITYYEGKHNHVPAA 467
Db 388 GYRWKYGQKVGKGNPNPNRSYKCTHPGCPVRKHVERASHDIRSVITYYEGKHNHVPAA 447
Qy 468 RGSQSHSVNPNPNPNPNASNHTNTAATSVRLLPVTHQSDNSIQNQ-----RSQAPPEG 519
Db 448 RGSQSHGLNR-----GANPNNAAMAMAIRP-----STMSLQSNYPIPISTRPQQEGQ 498
Qy 520 SPFTLEMLQPGSGFGSGGNPMQSYVYNQOQLSDNVFSSRTKEPRDDMLESL 574
Db 499 APY--EMLQSGGGYSGFGGNPMNAYAN--QIQDNAP--SRAKSEPRDDLFLDILL 548

```

RESULT 2

```

Q941B6 PRELIMINARY; PRT; 559 AA.
ID Q941B6
AC Q941B6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE WRKY DNA-binding protein.
GN NWRKY-6.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Petit Havana SRI;
RA Mao K., Hayashi S., Kojima-Suzuki H., Morikami A., Nakamura K.;
RT "Role of conserved residues of the WRKY domain in the DNA-binding
RT activity of tobacco WRKY family proteins.";
RL Submitted (Jun-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB063573; BAB61053.1; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW DNA-binding
SQ SEQUENCE 559 AA; 62288 MW; AODD3549C811B5AE CRC64;

Query Match 45.6%; Score 1389.5; DB 10; Length 559;
Best Local Similarity 51.6%; Pred. No. 1.2e-89;
Matches 316; Conservative 70; Mismatches 123; Indels 105; Gaps 23;

Qy 1 MASSSGSLDTSASANSFT-NFTFTSTHPTMTYFSDLLASPLDNNKPPQGG----- 49
Db 1 MASSGGNMNT--FMNSFTSNYSFS-----SFSDDLSDNNNNNNNNNNNSMNQKN 51
Qy 50 -----LSERTGSGVPRFKSTPPPSLPL--SPPISSPSYFAIPGLSPABELLSPV 98

```

```

Db 52 LNWGPDQRMHQKNDKVEPKFKSPCLPMTSSSPASPSSYLAPPSLSPSVLLDSPV 111
Qy 99 LNSSNLLSPPTGAFVAQSPFWKSSGNGNQIVKEEDKSFNFQFQTRSGPPASSTATY 158
Db 112 LFNNSNLLSPPTGSP-----GNLN--SKEDDSRISDFSFQSR-----ATSSSMF 155
Qy 159 QSSN-----VTVOTQOPWFSFOBATKQDNFSSGKGMKMTENSSSMQSPFSEIASVQTN 210
Db 156 QSSAPRNSLEDLWTRQ-----QHANOQNEFSTAKTTGVKSEVPIQSFQSE--KMQSNPPP 209
Qy 211 -HSNGFQSDYGNYPPOSQTLSSRSDGYNWRKYQKQVKGSENPRESYKCTPNCPTKK 269
Db 210 VHYT-----QPSQYVREQKAEQDYNWRKYQKQVKGSENPRESYKCTPNCPTKK 260
Qy 270 VERSLDGQITEIVYKGTNHPKPNTRRNSNS--SSLAIPHNSIRTEIPDQSYATHGSG 328
Db 261 VERNLDGHTITEIVYKGTNHPKPNTRRNSNS--SSLAIPHNSIRTEIPDQSYATHGSG 315
Qy 329 QMDSAATPENSISIGDDDFEQSSQKCKGGDEYDEDEPAKWKTEGENEGMSAPGSR 388
Db 316 QRSDFAGTNSASFGDEDIDQSPVSKGED--DQNEPEAKWKGDNEVEISS--ASRT 372
Qy 389 VREPRVVVQTTSDIDLDDGYRWKYGQKVGKGNPNPNRSYKCTHPGCPVRKHVERASHD 448
Db 373 VREPRVVVQTTSDIDLDDGYRWKYGQKVGKGNPNPNRSYKCTHPGCPVRKHVERASHD 432
Qy 449 LRAVITYYEGKHNHVPAAARGSGSHSVNPNPNPNASNHTNTAATSVRLLPVTHQ 504
Db 433 LRAVITYYEGKHNHVPAAARGSGSHSVNPNPNPNASNHTNTAATSVRLLPVTHQ 482
Qy 505 NSLQ-----NORSQAPPEGQSPFTLEMLQSGSGFGSGGNPMQSYVYNQOQLSDNV 555
Db 483 NSMQGNFNDTFNTRVQT--TQNPPTITLQMLQSSGSSSYSGPDTSSGSMNQMSNI 541
Qy 556 FSSRTKEPRDDMF 569
Db 542 -KPIITKEPKDDPF 554

RESULT 3
O81639 PRELIMINARY; PRT; 515 AA.
ID O81639
AC O81639;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Zinc finger protein.
GN ZFP1.
OS Pimpinella brachycarpa.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Pimpinella.
OX NCBI_TaxID=45043;
RN [1]
RP SEQUENCE FROM N.A.
RA Cho J.-I., Lee K.-W.;
RT "Pimpinella brachycarpa zinc finger protein PbZFP1 (WRKY1) mRNA.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF080595; AAC31956.1; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
SQ SEQUENCE 515 AA; 57553 MW; 8BF05537C621B46A CRC64;

Query Match 45.2%; Score 1379.5; DB 10; Length 515;
Best Local Similarity 55.1%; Pred. No. 5.7e-89;
Matches 313; Conservative 57; Mismatches 123; Indels 75; Gaps 24;

Qy 28 MTTSFSDLLASPLDNNKPPQGGISERTGSGVPRFKSTPPPSLPLSPPISSPSYFAIPPG 87
Db 1 MTSSLDLLAQPNDDFGSNWGFENQ-----KTKSFANPQSPFPSPSPSYFS----- 50

```



```

QY 88 LSPAEILLDSPVLNLSNLLPPTTGAFVAQSFNWKSSGGNQOIIVKEEDKSFNSFQTR 147
Db 51 -----FLDSPIQNNYIVSSSGNGTMAQSF-----EENQNFSDSFPTQ 92
QY 148 SGPASSTATYQSSNVTV-----QOPWSFQEQATKQDNFSSGGKMKMTENSSMS 197
Db 93 S-RPASSTSS--SSFPANTNSVEESLKRKQGGWNPFEAAKNEFORFSPFM-TMNQANM 148
QY 198 QGSPPIASVQTN-HSG--FOSDYGNYPPQSOTL-----SRRSDGYNWRKYGQKQVKG 249
Db 149 --LSPMTWQANMQSNAVALQNSINYAQSSOSSQTNRDQSKLDDGYNRKYGQKQVKG 206
QY 250 SENPRSYKCTYPCNCTKKKVERSLDQITEIVYKGTNNHPKPQNTNRNNSSSSLAIPH 309
Db 207 SENPRSYKCTYPCNCTKKKVERSLDQITEIVYKGTNNHPKPQNTNRNNSSSSLAIPH 309
QY 310 SNSIRTEIPQSVATHSGQMSAATPENSISIGDD--FEQSSQCKSGGDEYDEDEPD 368
Db 261 QNSIPT-MPETSLL--NGHLEPVTTTPENSSLSFGEDDLFEQSSMN-KQGD--DENEFD 314
QY 369 AKRWKIEG-NEGMSAPGSRVREPRVVQTTSDIDLDGGRWRKYGQKVGKGNPNRS 427
Db 315 AKRWKEYENNETMSSLSGRTVREPRVVQTTSDIDLDGGRWRKYGQKVGKGNPNRS 374
QY 428 YKCTHGPCVPRKHVERASHDLRAVITYEGKHNHDVPAARGSGSHSVNRPMPNNAHNT 487
Db 375 YKCTHGPCVPRKHVERASHDLRAVITYEGKHNHDVPAARGSGSVAVNRPSDNTA--T 431
QY 488 NTAATSVRLLPVTHQSDNSLQNSORQAPEQSGPFTLEMLQSPGSGFGSGFQGNPMQSY-V 546
Db 432 TSAPTATR--PTTNYL-NPLQNRAQ--PANGQAPFTLEMLQSPGSGFGSGFQGNPMQSY 487
QY 547 NQOQLSDNVFSRTKEPRDDMELESIL 574
Db 488 NQOQASGQFST-AKDEPDVDSFDFSFL 514

```

RESULT 4

```

Q9SQ04 ID Q9SQ04 PRELIMINARY; PRT; 514 AA.
AC Q9SQ04;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Zinc-finger type transcription factor WRKY1.
GN WRKY1.
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Petroselinum.
OX NCBI_TaxID=4043;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400614; PubMed=10469648;
RA Eulgem T., Rushton P.J., Schmelzer E., Hahlbrock K., Somssich I.E.;
RT "Early nuclear events in plant defence signalling: rapid gene
RT activation by WRKY transcription factors.";
RL EMBO J. 18:4689-4699(1999).
DR EMBL; AF121353; AAD5974.1; -
DR InterPro; IPR001360; Glyco_hydro_1.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00811; WRKY.
SQ SEQUENCE 514 AA; 57321 MW; 279064A5FF85D8D5 CRC64;

```

Query Match 44.0%; Score 1342; DB 10; Length 514;
 Best Local Similarity 52.2%; Pred. No. 2.5e-86;
 Matches 236; Conservative 74; Mismatches 123; Indels 74; Gaps 22;

```

QY 28 MTTFSDDLASPLDNNKPPQGLSERTGSGVPRFKSTPPPLSPSPSSYFAIPPG 87
Db 1 MSSSLGDLAQPNNDDFGSNWGFENQ-----NFKSFANGSLFPSPSPSSYFVS---- 50

```

```

QY 88 LSPAEILLDSPVLNLSNLLPPTTGAFVAQSFNWKSSGGNQOIIVKEEDKSFNSFQTR 147
Db 51 -----FLDSPIQNNYIVSSSGNGTMAQSF-----EENQNFSDSFPTQ 92
QY 148 SGPASSTATYQSSNVTV-----QOPWSFQEQATKQDNFSSGGKMKMTENSSMS 201
Db 93 TRPASSTSSSFPANTNLNVEDSLKRKQGGWNPFEAKNDFLMDNASV-TSDIATLQRTS 151
QY 202 PEIASVQTNHNS-----GFQSDYGNVPPQSOTL-----SRRSDGYNWRKYGQKQVKGSE 251
Db 152 PMWT--MNHANMQSNAVALQNSINYAQSSOSSQTNRDQSKLDDGYNRKYGQKQVKGSE 208
QY 252 SENPRSYKCTYPCNCTKKKVERSLDQITEIVYKGTNNHPKPQNTNRNNSSSSLAIPH 311
Db 209 SENPRSYKCTYPCNCTKKKVERSLDQITEIVYKGTNNHPKPQNTNRNNSSSSLAIPH 311
QY 312 SIRTETIPQSVATHSGQMSAATPENSISIGDD--FEQSSQCKSGGDEYDEDEPD 370
Db 263 SIGT-MPESLLE--NGRSEPVTTTPENSSLSFGEDDLFEQSSMN-KQGD--DENEPSDK 316
QY 371 RWKIEG-NEGMSAPGSRVREPRVVQTTSDIDLDGGRWRKYGQKVGKGNPNRSY 429
Db 317 RWKIEG-NEGMSAPGSRVREPRVVQTTSDIDLDGGRWRKYGQKVGKGNPNRSY 376
QY 430 KCTHGPCVPRKHVERASHDLRAVITYEGKHNHDVPAARGSGSH-SVNRPMNNAHNT 488
Db 377 KCTHGPCVPRKHVERASHDLRAVITYEGKHNHDVPAARGSGSVAVNRPSDNTA--T 431
QY 489 NTAATSVRLLPVTHQSDNSLQNSORQAPEQSGPFTLEMLQSPGSGFGSGFQGNPMQSY-V 547
Db 432 TSAPTATR--PTTNYL-NPLQNRAQ--PANGQAPFTLEMLQSPGSGFGSGFQGNPMQSY 487
QY 548 NQOQLSDNVFSRTKEPRDDMELESIL 574
Db 488 NQOQASGQFST-AKDEPDVDSFDFSFL 514

```

RESULT 5

```

Q40827 ID Q40827 PRELIMINARY; PRT; 514 AA.
AC Q40827;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE WRKY1.
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Petroselinum.
OX NCBI_TaxID=4043;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051827; PubMed=8896462;
RA Rushton P.J., Torres J.T., Parniske M., Wernert P., Hahlbrock K.,
RA Somssich I.E.;
RT "Interaction of elicitor-induced DNA-binding proteins with elicitor
RT response elements in the promoters of parsley PrL genes.";
RL EMBO J. 15:5690-5700(1996).
DR EMBL; U48831; AAC49527.1; -
DR TRANSFAC; T03718; -
DR InterPro; IPR001360; Glyco_hydro_1.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00811; WRKY.
SQ SEQUENCE 514 AA; 57206 MW; 3B03C8A7DA486638 CRC64;

```

Query Match 43.6%; Score 1331; DB 10; Length 514;
 Best Local Similarity 52.5%; Pred. No. 1.5e-85;
 Matches 296; Conservative 72; Mismatches 128; Indels 68; Gaps 22;
 QY 28 MTTFSDDLASPLDNNKPPQGLSERTGSGVPRFKSTPPPLSPSPSSYFAIPPG 87

```
Db 1 MSSLGLDLAQNNDDEGSGWGFENQ-----KTKSPANQSLPSPVPPVSPSSYFS----- 50
Qy 88 LSPAEILDSPVLNSSLNIPPTTGAFVAQSFNWKSSGGNQOIVKEEDKSFNSFQTR 147
Db 51 -----FLDSPITQNNNYTIVSSSGNGTMAQSF-----KEENQNFDSFSPAQ 92
Qy 148 SGPPASSTATYQSSNVTV-----OTQOPWFOEATKQDNFSSGKGMKTNSSSSMSQFS 201
Db 93 TRPASSTYSSSLIPANTNLVEDSLKRGKQGMWFEPAPKNDPLMDNASV-TSDIATLORIS 151
Qy 202 PEIASVQTN-HSN-GFOSDYGYNPPQSTL-----SRRSDGYNWKYQKQVKGSENPR 254
Db 152 PEMTMQANMQSNAALQSNLNNYAQSQSQTNRDQSKLDGYNWKYQKQVKGSENPR 211
Qy 255 SYIKCTYPCPTKKKVERSLDQITVYKGTHTNHPKQNTNRNSSLAIAPHNSIR 314
Db 212 SYIKCTYPCPTKKKVERSLDQITVYKGTHTNHPKQNTNRNSSLAIAPHNSIR 314
Qy 315 TEIPDQSVATHGSGQMDSAATPENSISIGDD-PEOSQKCKGGDEYDEPDARWK 373
Db 266 T-MPESSLE--NGRSPVTPPENSISLFGEDDLFEGGSMY-KGDD--DGNEDPSRWK 319
Qy 374 IEGE-NEGMSAPGSRTRVPRVVQTTSDIDILDDGYRWKYGKQVKGPNPRSYIKCT 432
Db 320 GEYESNEPMSLSGSRTRVPRVVQTTSDIDILDDGYRWKYGKQVKGPNPRSYIKCT 379
Qy 433 HPGCPVRKHVRASHDLRAVITTYEGKHNHDVPAARGSGH-SVNRPMNNAASHHTAA 491
Db 380 QVGCPCVRKHVRASHDLRAVITTYEGKHNHDVPAARGSGH-SVNRPMNNAASHHTAA 491
Qy 492 TSVRLLPVTHOSNLSNORSOAPEGSPPTLEMLQSPGSGFGSGNPMQSY-VNQOQ 550
Db 435 TAIR--PTTNVL-NPLQNPRAQ-PANGQAPPTLEMLQSPGSGFGSGNPMQSY-VNQOQ 490
Qy 551 LSDNVFSRTKEPRDDMFLESLL 574
Db 491 QASGQFST-AKDEPDVDSFDSFL 513
```

RESULT 6

```
Q9XJ26 ID Q9XJ26 PRELIMINARY; PRT; 477 AA.
AC Q9XJ26;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE NtWKY1.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Xanthi;
RA Shinghi H., Yamamoto S., Suzuki K.;
RT "Analyses of an elicitor-responsive element and transcription factors
in cultured tobacco cells."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB022693; BAA82107.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY. 2.
DR PROSITE; PS0811; WRKY. 2.
SQ SEQUENCE 477 AA; 53053 MW; 24B8617DB8C59B7A9 CRC64;
```

```
Query Match 41.9%; Score 1278; DB 10; Length 477;
Best Local Similarity 53.9%; Pred. No. 7.1e-82;
Matches 279; Conservative 64; Mismatches 109; Indels 66; Gaps 17;

Qy 71 LSPPPIPSSSYFAIPGGLSPAELDSDPVLNSSLNIPPTTGAFVAQSFNWKSSGGNQ 130
Db 2 ISSSPASPSSYLAFPPLSPVLLDSPVLFNNSNLTSPPTGTF-----GNLN 49
```

```
Qy 131 IVKEEDKSFNSFQTRSGPPASSTATYQSSN-----VTVOTQPMWFOEATKQDNFSSG 185
Db 50 -SKEDNSRLSDFSFQSKA---ATSSSMFQSSAPRNSLEDLMTQK---QHANQNEPSTV 101
Qy 186 KGMKMTENSSMSQSPSPETIASVOTNHSNGFQSYGNYPPOSQTLRRSDGYNWKYQK 245
Db 102 KTRGVKSEVAPIQSFQSE-----KMQSNPAPVHYTH--PSQYVREQKAEDGYNWKYQK 154
Qy 246 QVKGSENPRSYIKCTYPCPTKKKVERSLDQITVYKGTHTNHPKQNTNRNSSLN-SS 304
Db 155 QVKGSENPRSYIKCTYPCPTKKKVERSLDQITVYKGTHTNHPKQNTNRNSSLN-SS 304
Qy 305 LAIPHNSIRTEIPDQSVATHGSGQMDSAATPENSISIGDDPEOSQKCKGGDEYDE 364
Db 215 LA--YSN--LDITNQSNFLDNAQRDSFAGTDSNSSFQDEIDQGSPIKSGED--DG 267
Qy 365 DEPDAKRWKIEGNEGMSAPGSRTRVPRVVQTTSDIDILDDGYRWKYGKQVKGPN 424
Db 268 NEPEPKRWKGDNEVEISS-ASRTVREPRIVVQTTSDIDILDDGYRWKYGKQVKGPN 326
Qy 425 PRSYIKCTHPCVRKHVRASHDLRAVITTYEGKHNHDVPAARGSGHSHSVNRPMNNA 484
Db 327 PRSYIKCTHPCVRKHVRASHDLRAVITTYEGKHNHDVPAARGSGHSHSVNRPMNNA 386
Qy 485 NHTNTAATSVRLLPVIHQ-----SDNSLQ-----NORSQAPPEGQSPFTLEMLQSPG 531
Db 387 NS-----MPVPRPSMLANNNOGMNFNTFFNTRVQT-TQNPPTITLQMLQSSG 435
Qy 532 SFQSGFGNPMQSYVNVQQLSDNVFSRTKEPRDDMF 569
Db 436 NSSYSGFDTSYGSGYNNQMPMNNI-KPISKEEPKDDFF 472
```

RESULT 7

```
Q9SXP5 ID Q9SXP5 PRELIMINARY; PRT; 353 AA.
AC Q9SXP5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Transcription factor NtWKY2.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Xanthi;
RA Shinghi H., Yamamoto S., Suzuki K.;
RT "Analyses of an elicitor-responsive element and transcription factors
in cultured tobacco cells."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB020590; BAA77383.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY. 2.
DR PROSITE; PS0811; WRKY. 2.
SQ SEQUENCE 353 AA; 40061 MW; 33E6D3A1FF21317E CRC64;
```

```
Query Match 40.1%; Score 1222; DB 10; Length 353;
Best Local Similarity 65.3%; Pred. No. 4.1e-78;
Matches 245; Conservative 42; Mismatches 60; Indels 28; Gaps 10;
```

```
Qy 207 VQTNHNGFOSDYGNTPPQSQTLRRSDGYNWKYQKQVKGSENPRSYIKCTYPCN 266
Db 1 MQSLKXNGQSNQYNNQSSQSIREQKRSDDGYNWKYQKQVKGSENPRSYIKCTYPCN 60
Qy 267 KKXVERSLDQITVYKGTHTNHPKQNTNRNSSLNLAIP---HNSNIRTEIPDQSYA 323
Db 61 KKXVERSLDQITVYKGTHTNHPKQNTNRNSSLNLAIP---HNSNIRTEIPDQSYA 323
Qy 324 THSGQGMDSAAATPENSISIGDD--DPEQSKCKGGDEYDEPDARWKIEGENEGM 381
```

```

Db 117 SYGNS-----ATPENSISFCDHDEQSSQSRSGDDNBEEDPSKRWKRESEGL 171
Qy 382 SAGSRTVRPRVVVQTTSIDILDDGYRWKYGKQVKNPNPRSYKCTHPCGCPVRKH 441
Db 172 SAGSRTVRPRVVVQTTSIDILDDGYRWKYGKQVKNPNPRGYKCTSGCPCVRKH 231
Qy 442 VERASHDLRAVITTYEGKHNDVPAARGSGSHSVNRPMPNASHNTAATSVRLLPVH 501
Db 232 VERASQDIRSVITTYEGKHNDVPAARGSG---INRPVAFNIT--YNGGANAMAIRPSV- 285
Qy 502 QSDNSLQNSQAPPEGQSPTLEMLQSPGF-GFSGFGNPMQSYVNOQLSDNVFSSRT 560
Db 286 TSQIPLOSIRPQ-----QSPFTLEMLKHPNSYNGFGSVNSEDSEYEN--QLQDNVFSRA 338
Qy 561 KEPRDDMFLESLLC 575
Db 339 KDEPRDDMFMETLLC 353

RESULT 8
Q941B4 PRELIMINARY; PRT; 378 AA.
AC Q941B4; 2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE WRKY DNA-binding protein (Fragment).
GN NTRKY-8.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Petit Havana SRI;
RA Maeo K., Hayashi S., Kojima-Suzuki H., Morikami A., Nakamura K.;
RT "Role of conserved residues of the WRKY domain in the DNA-binding
RT activity of tobacco WRKY family proteins."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB063575; BAB61055.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW DNA-binding.
FT NON TER.
SQ SEQUENCE 378 AA; 43050 MW; DA0BC909C078413D CRC64;

Query Match 39.4%; Score 1201.5; DB 10; Length 378;
Best Local Similarity 56.8%; Pred. No. 1.3e-76;
Matches 254; Conservative 48; Mismatches 58; Indels 77; Gaps 15;

```

```

Qy 137 KPSFNFSQTRSGPPASSTAYSSNVTVQTPQSPQFQATKQDNFSGGKMKNTSSS 196
Db 1 KNYSDFSFQ---PQFASVSPSQTNVPVL-GKQSNYQESRKQND-----ENAG 45
Qy 197 MQSPSPFIASVQTHNSGFQSDYGNYPQSQTLRRSDGYNWRKYQKQVKGSENP 256
Db 46 -----TSELQSLKNGQSNQYNKQ-----SSRSEDGYNWRKYQKQVKGSENP 91
Qy 257 YKCTYPCNPTKKKVERSLDQGITIVYKGTNHPKP-QNTRRNSSNSLSLAIPHNSIRT 315
Db 92 YKCTFPNCTPKKVERCLDQGITIVYKGTNHPKPQTQSTR-----SSLSLAIPYNTQTN 147
Qy 316 EIPQSYATHSGQMDSAATPENSISIGDDDFEQSQKCKSGGDEYDED-----EPDA 369
Db 148 EIPDHQ-----STPENSISIFGDDHEKS-----RSRGDDFDEEEDPDP 191
Qy 370 KRWKIEGEMGSAAGSRTVRPRVVVQTTSIDILDDGYRWKYGKQVKNPNPRSY 429
Db 192 KRWKRESESEGLSAPGSGRTVRPRVVVQTTSIDILDDGYRWKYGKQVKNPNPRSY 251

```

```

Qy 430 KCTHPCGCPVRKHVERASHDLRAVITTYEGKHNDVPAARGSGSHSVNRPMPNASHNT 489
Db 252 KCTSPGCPVRKHVERASQDIRSVITTYEGKHNDVPAARGSG---AINRPVA-----PTIT 303
Qy 490 KATSVRELLPVTHQSDNSLQNSQAPPEGQSPTLEMLQSPGF-GFSGFGNPMQSYVNO 548
Db 304 YNNAIPRPSV-----TSQIPLOQSPPTLEMLKHPNSYNGFGSVNSEDSEYEN- 352
Qy 549 QQLSDNVFSSRTKEPRDDMFLESLLC 575
Db 353 -QLQDNVFSRAKDEPRDDMFMETLLC 378

RESULT 9
Q888P5 PRELIMINARY; PRT; 512 AA.
AC Q888P5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative WRKY-type DNA binding protein.
GN AT2G38470.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Roundley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Town C.D., Kaul S.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004683; AAM14994.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
SQ SEQUENCE 512 AA; 56457 MW; 8F19CBE41BC18662 CRC64;

Query Match 37.4%; Score 1139.5; DB 10; Length 512;
Best Local Similarity 44.7%; Pred. No. 4.4e-72;
Matches 254; Conservative 80; Mismatches 144; Indels 90; Gaps 15;

```

```

Qy 40 LDNNKPPQ-----GGLSERTGS-----GYPKFKSTPPPLPLSPPISSSYPAIP 85
Db 1 MDNSRTQNMNGSANWSQSGRTSTSSLEDLEIPKFRSPAPSSISISPLSVSTCF- 57
Qy 86 PGLSPAEILDSPVLLNSS-NILPSPTTGAPVAOSFNWSSSGGNQOIVKEEDKSFN- 141
Db 58 ---SPSLFLDSPAFVSSSANVLASPTTGALI-----TNVTNQKINEGDKSNNNFN 106
Qy 142 ---FSQTSRGPASSTATYQSSNVTVQTPQSPQFQATKQDNFSGGKMKNTSSSMQ 198
Db 107 LDFPSFHTQSGVSAPTTTTTTTTTTTTTNS--SIFSQEQQKKQNSEQWSTETRPNQ 164
Qy 199 SFSPEIASVQTNHNSGFQSDYGNYPQSQTLRRSDGYNWRKYQKQVKGSENP 258
Db 165 AVS-----YNGRE-----QRKGEDGYNWRKYQKQVKGSENP 201
Qy 259 CTYPNCPCTKKKVERSLDQGITIVYKGTNHPKPQNTRRNSSNSLSL-----AIPHSN 311
Db 202 -CTFPNCPCTKKKVERSLDQGITIVYKGTNHPKPQNTRRNSSNSLSL-----AIPHSN 261

```

```

QY 312 SIRTEIPDQSYATHGSGQMSAATPENSISIGDDDFEQSSQKCKSGGDEYDEPDQAKR 371
Db 262 QASDQPNNSNFHQSDSGFQMQEDNTTSDSGDDEFEQGS-SIVRDEDDCCSEPEAKR 320
QY 372 WKIEGENEGMSAPGSRVREPRVVVQTTSDIDLDGGRWKYQKQVKGKGNPNRPSYK 431
Db 321 WKGDNETNGGGGSKTVREPRVVQTTSDIDLDGGRWKYQKQVKGKGNPNRPSYK 380
QY 432 THPGCPVKHVERASHDLRAVITTYEGKHNDVPAARGSGSHSVNRPMPNPNASHTNTAA 491
Db 381 TTIGCPVKHVERASHDLRAVITTYEGKHNDVPAARGSG-YATNRAPOQSSSVPIRPA 439
QY 492 TSVRLLPVTHQSDNSLQNSQAPPEQSPFTLEMLQ---SPGSGFGSGFGNPMQSYVN 547
Db 440 IA-----GHSNYTSSQAPYTLQMLHNNNTNTGPGFYAMNNNNNSLQ 483
QY 548 QQQLSDNVFSRTKEPRDD-MFLESLL 574
Db 484 TQNFVGGGFSRAKEEPEETSFFDSFM 511

```

RESULT 10

```

Q9XFF1 ID Q9XFF1 PRELIMINARY; PRT; 501 AA.
AC Q9XFF1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE DNA-binding protein WRKY1.
GN WRKY1.
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Aveneae; Avena.
OX NCBI_TaxID=4498;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Rhiannon; TISSUE=aleurone;
RA Screen S.E., Powell R., Rollason R., Huttly A.K., Lazarus C.M.;
RT "Isolation of WRKY-type DNA-binding proteins from Avena sativa.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF140554; AAD32677.1; -.
DR InterPro: IPR003657; WRKY.
DR Pfam: PF03106; WRKY; 2.
DR PROSITE: PS50811; WRKY; 2.
KW DNA-binding.
SQ SEQUENCE 501 AA; 53816 MW; E9C74BE8ED21E4D8 CRC64;

```

```

Query Match 36.6%; Score 1116.5; DB 10; Length 501;
Best Local Similarity 46.0%; Pred. No. 1.8e-70;
Matches 256; Conservative 60; Mismatches 135; Indels 105; Gaps 19;

QY 64 TPPSLPLSPPTSPSSYFAIPPLGLSPAELLDSVLLNSNLPSPPTTGAFVAQSFNW-- 121
Db 4 TPTTSP----PPASPSSYF-----NNMSTGFLDSILLTPS-LFSPPTGFPPLPLNMWG 54
QY 122 ---KSSGGGNOQIVKE-EDKSFNFSQTRSGP-PASSTAT-----YQSS----- 161
Db 55 TAPESNDGLQLGSKVDGQQRQYSGFTQTAAAPVPAAMPNTNTTASSFLQSSMPMAQQGH 114
QY 162 -NVTVTQOPKSPFQATKQDNFSGKGMKMTENSQSPSPETASVQTNHNSGFQSDYG 220
Db 115 DSYTGEQQPWSYQDAGSMD-----AMTTRPASFTPYEAPDMVG-NGGYNNAAPVSSG 167
QY 221 NYPQSQOTLSRR--SDDGYNWKYQKQVKGKGNPNRPSYKCTYPCNCTPKKKVERSLDQ 278
Db 168 TTAGYGRVQRRPSSDDGYNWKYQKQKMGKGNPNRPSYKCSFAGCSTKKKVEQAPDGV 227
QY 279 TEIVYKGTNHHPKQNTNRNSNSSSLAIPHSNSIRTEIPDQSYATHGSGQMSAATP 338
Db 228 TEIVYKGTNHHPKQNTNRNSAPASS-----SYASPASSDALLSGTPEN 271

```

```

QY 339 SSISIGDDDFEQSSQ---KCKSGGDEYDEPDQAKRWKIEGENEGMSAPGSRVREPRV 394
Db 272 SSASYGDDDETNGVSSALACQFGGGEFADNEPDSKRWRTDSDAEGVPVGANRTVREPRV 331
QY 395 VVQTTSDIDLDGGRWKYQKQVKGKGNPNRPSYKCTHPCGPKHVERASHDLRAVIT 454
Db 332 VVQTTMSDIDLDGGRWKYQKQVKGKGNPNRPSYKCTTAGCPVKHVERASHDLRAVIT 391
QY 455 TYEGKHNDVPAARGSGSHSVNRPMPNPNASHTNTAATSVRLLPVTHQSDNSLQ----- 508
Db 392 TYEGKHNDVPAALRGSA-----AAARYRAAPM--QAASYLQGGGGY 432
QY 509 -----NQRQAPPEQSPFTLEMLQSPGSGFGSGFGNPMQSYVNQOQLSDNVF--SS 558
Db 433 SSLRPDGGGGAQAPADQGFALS-----GFD-YNNSSYSYSGMQQNDAMYDAA 483
QY 559 RTKEPRDDMFLESLL 574
Db 484 RTKDEPRDDMFESQL 499

RESULT 11
Q94D89 ID Q94D89 PRELIMINARY; PRT; 357 AA.
AC Q94D89;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative DNA-binding protein ABF1.
GN P0439E11.17.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0439E11.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003315; BAB61266.1; -.
DR Gramene: Q94D89; -.
DR InterPro: IPR003657; WRKY.
DR Pfam: PF03106; WRKY; 2.
DR PROSITE: PS50811; WRKY; 2.
KW DNA-binding.
SQ SEQUENCE 357 AA; 38867 MW; E23FE44F5B550854 CRC64;

Query Match 35.4%; Score 1081; DB 10; Length 357;
Best Local Similarity 59.4%; Pred. No. 3.5e-68;
Matches 227; Conservative 28; Mismatches 73; Indels 54; Gaps 11;

QY 216 QSDYGNVYPOSQTLRRSDDGYNWKYQKQVKGKGNPNRPSYKCTYPCNCTPKKKVERSLD 275
Db 6 QGGGGYSQSQSQ--RSDSDGYNWKYQKQVKGKGNPNRPSYKCTYPCNCTPKKKVERSLD 63
QY 276 GQTEIVYKGTNHHPKQNTNRNSNSSSLAIPHSNSIRTEIPDQSYATHGSGQMSAAT 335
Db 64 GQTEIVYKGTNHHPKQNTNRNSNSSSLAIPHSNSIRTEIPDQSYATHGSGQMSAAT 115
QY 336 PENSSISIGDDDFEQSSQK--KSGGDEYDEPDQAKRWKIEGENEGMSAPGSRVREPRV 394
Db 116 PENSSASFGDDEIRVSGPRAGNGGGDEFFDDPSKRWKIDGGEGISMAGNRTVREPRV 175
QY 395 VVQTTSDIDLDGGRWKYQKQVKGKGNPNRPSYKCTHPCGPKHVERASHDLRAVIT 454
Db 176 VVQTTMSDIDLDGGRWKYQKQVKGKGNPNRPSYKCTTAGCPVKHVERASHDLRAVIT 235
QY 455 TYEGKHNDVPAARGSGSHSVNRPMPNPNASHTNTAATSVRLLPVTHQSDNSLQ--RS 512
Db 236 TYEGKHNDVPAARGSA--ALYRPAAPAA-----AATS-----SHFPLNPQPPM 278

```

```
Qy 513 QAPPEGQSPFTLEMLQSPGSGF-----GFGSGF-----GNPMQSYVNOQOOLS 552
Db 279 SYQPTGQPVYALR-----PDGFGGQPFPGVVGSGFGSGFGFDDARGSYMSHQOQOQRN 334
Qy 553 DNVFSSRTKEPRDDMFLESLL 574
Db 335 DAMHASRAKEPDDMFQNSL 356

RESULT 12
Q38704 PRELIMINARY; PRT; 402 AA.
ID Q38704
AC Q38704;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE DNA-binding protein (Fragment).
OS Avena fatua.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Aveneae; Avena.
OX NCBI_TaxID=4499;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA; TISSUE=Aleurone;
RX MEDLINE=96128012; PubMed=8541496;
RA Rushton P.J., Macdonald H., Huttly A.K., Lazarus C.M., Hookey R.;
RT "Members of a new family of DNA-binding proteins bind to a conserved
RT cis-element in the promoters of alpha-amyl 2 genes.";
RL Plant Mol. Biol. 29:691-702(1995).
DR EMBL; Z48429; CAA88326.1; -
DR TRANSFAC; T03726; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW DNA-binding.
FT NON TER
SQ SEQUENCE 402 AA; 43232 MW; DA81C09530E9A868 CRC64;

Query Match 33.98; Score 1032.5; DB 10; Length 402;
Best Local Similarity 53.2; Pred. No. 1.1e-64;
Matches 225; Conservative 48; Mismatches 105; Indels 45; Gaps 15;

Qy 168 QPMFQBEATKQDNFSSGKGMKMTSSMSQSFPEIASVQTNHNSGFQDYGNYPPQSQ 227
Db 2 QQPMGYQQ-----PGAM--DAGANAASF--APAVQATSS--MAPSGGVTRQTH 46
Qy 228 TLRRSDGYNWRKYQKQKVGSENPRSYKCTYPNCTPKKVERS LDGQITEIVYKGT 287
Db 47 SQRSSDDGYNWRKYQKQKVGSENPRSYKCTFPNCTPKKVERSIEGQITEIVYKGT 106
Qy 288 NHPKPQNTRRNSSSSSLAIPHNSIRTEIPDQSYATHSGQMDSA--ATPENSISIGD 345
Db 107 NHAKPLSTRSGGGGGGGA-----QVLSQGGGGDASEHSGAMGAPVSTPENSASF 162
Qy 346 DDFEQSQKCKS--GGDYDEDPDAKRWKTEGENEGM--SAPGRTVREPRVVVQTTSDID 403
Db 163 DETGASSPRAGNVGGDDLDDEPDSSKWRKDGDEGSGNSVAGNTVREPRVVVQTTSDID 222
Qy 404 ILDDGYRWKYQKQKVGKGNPNRPSYKCTHPGCPVRKVERASHDLRAVITTYEGKHNH 463
Db 223 ILDDGYRWKYQKQKVGKGNPNRPSYKCTTVGCPVRKVERASHDLRAVITTYEGKHNH 282
Qy 464 VPAARGSGSHSVNRPMPNNSN-----HTNTAATSVRLLP--VIHQSDNSLQNSQAPPEG 518
Db 283 VPAARGSA--ALYRPAAPRAAADSAMSTSQQYTANQQQPSAMTYQTSAAGTQQYAPRPG 340
Qy 519 -----QSPFTLEMLQSPGSGFGS--GFGNPMQSYV-----QQQSDNVFSSRTKEEPRDDM 568
Db 341 FGQSQGSGFGFN-----GSGFGSAGFNDPTGYSYMSHQOQOQRNDAMQASGAKEEPRDM 395
Qy 569 FLE 571
```

```
Db 396 FFQ 398

RESULT 13
Q94IB3 PRELIMINARY; PRT; 727 AA.
ID Q94IB3
AC Q94IB3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE WRKY DNA-binding protein.
GN NTRKY-9.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Petit Havana SRI;
RA Maao K., Hayashi S., Kojima-Suzuki H., Morikami A., Nakamura K.;
RT "Role of conserved residues of the WRKY domain in the DNA-binding
RT activity of tobacco WRKY family proteins.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB063576; BAB61056.1; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW DNA-binding.
SQ SEQUENCE 727 AA; 78975 MW; A5674A4F282F6F24 CRC64;

Query Match 30.4%; Score 927; DB 10; Length 727;
Best Local Similarity 38.1%; Pred. No. 6.3e-57;
Matches 237; Conservative 79; Mismatches 154; Indels 152; Gaps 23;

Qy 41 DNKKPPQGGGLSE-----RTGSGVKEF--KSTPPSLPLSPPISSPYFAIPGLSPAELL 94
Db 74 DQKMSRGGGLERMAARAGFNAPKLNTESLRPADLSQNGVRSF--YLTIPPLGLSPTSL 131
Qy 95 DSVILNSNLLPSPTTGAFVAGSFNWKSSGGNQIQIVEEDK-----SFSNPSF- 144
Db 132 ESPVFLSNLVSPTTGKFP--QFASGIESRNTFTMEDPDPRKKNALNESINSSSF 188
Qy 145 ---QTRSGPPASATATYQGSNVT-----VOTQOPWSFOE 175
Db 189 PVDETAPSLPFGTTSRVNSNTISQQCFPIKIVSVHSQNSLLSHSVATQMTQSEKGLHQ 248
Qy 176 ATKQDNFSSGKGMKMTSSMSQSFPEIASVQTNHNSGFQSDYGNYPPOSQTLRR--- 232
Db 249 SSDPFRFSAEKGVDRDSNVTPESRNFQ---SVGSNMEHSGPPLD---EPQDEIDQKRVGG 300
Qy 233 -----SDDGYNWRKYQKQKVGSENPRSYKCTYPNCTPKKVERS LDGQITEIVYK 284
Db 301 DPNVVGAPADGYNWRKYQKQKVGSEYPRSYKCTHPNCPVKKKVERSHGHTIETIYK 360
Qy 285 GTNHHPKPQNTRRN---SNS-----SSLAI PHNSNI--RTEI---PD----- 319
Db 361 GANHHPKPPNRRKRSALGSNLSLGLQLDGAQVQSVSGNSGLGRANTQKAPDAGGLDWRNN 420
Qy 320 -----QSYATHSGQMDS--AATPENSISISIGDDDFEQSQKCKSGGD 360
Db 421 NLDVTSAAHLGSAYCNCSASFVQNNVTQLESQGVADVSTFSFNSDENDEDDRGTHGSVQGY 480
Qy 361 EYDEDEPDAKRWKIEGENEGMSAPGRTVREPRVVVQTTSDIDILDDGYRWKYQKGVK 420
Db 481 DGEDESSESKRRKLETYSIDMSG--ATRAIREPRVVVQTTSEVDILDDGYRWKYQKGVK 539
Qy 421 GNPNPSYKCTHPGCPVRKVERASHDLRAVITTYEGKHNHDPVPAARGSGSHSVNRPMP 480
Db 540 GNPNPSYKCTSAGCNVRKVERASHDLKLSVITTYEGKHNHDPVPAARN----- 588
Qy 481 -NNASNHTNTAATSVRLLP-----IHQSDNS--LQNORSQAPPEGQSPFTLEMLQSPG 531
```

```

Db 589 ---SSHVNSASNTLPAPVTPAPQAQSHLRPEPAQLQANVAR-----PDRQPSLG 635
QY 532 SFGSGFGNPMQSY-VNQOQLS 552
Db 636 SFG-----PTGYSYGINQOGLA 653

RESULT 14
QY Q9ARC7 PRELIMINARY; PRT; 739 AA.
ID Q9ARC7
AC Q9ARC7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 79.9 kDa protein.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21178822; PubMed=11283350;
RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,
RA Schumacher K., Schmitt G., Schmidt R.;
RT "Comparative sequence analysis reveals extensive microcolinearity in
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella
RT genomes.";
RL Plant Cell 13:979-988(2001).
DR EMBL; AJ303345; CAC36402.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW Hypothetical protein.
SQ SEQUENCE 739 AA; 79854 MW; F9CAF86DA15B8722 CRC64;

Query Match 30.3%; Score 924; DB 10; Length 739;
Best Local Similarity 35.6%; Pred. No. 1.1e-56;
Matches 242; Conservative 95; Mismatches 177; Indels 166; Gaps 25;

QY 36 LASPLDNNKPPQGG-----LSERTGSGVPRF--KSTPPPSLPLSPPISSSYFAIPPLG 88
Db 78 LSPSPDQKMSRGGLLRIARAGFNAPKLNTESLRPAQLRQNGVRSP--YLTIPPLG 135
QY 89 SPALLDPSVLLNSNLPSTTCAFAVQS--FNWSSSGGNQIIVKEEDKSF-----SNF 142
Db 136 SPTLLSPVFLSNLVQSPSTTGKFLFSSGIDNRNSALMMDTDKRDNALESINSSSF 195
QY 143 SFQ-----TRS-----GPPASSTATYQSSNVTVQTO-----Q 169
Db 196 SFKVPETAPSLPFGTTTSSRLQVNPNSFQGGFPNIEVSHVSNLSLVRHEATQNPQT 255
QY 170 PWSFOEATKQDNFSSGKGMKNTENSSMQSFSPFIASVQTNHNGFQSDYGNYPQSQTL 229
Db 256 NGTLQSSDPRFSAEKDVM--ANNVLDSTFTQVGSADVHSPPLDE-----PQEDI 307
QY 230 SRR-----SDDGYNRWKYQKQVKGSENPRSYKCTYPCNCTPKKKVERSLDQI 278
Db 308 DQGGGDPNVAGAPAEADGYNRWKYQKQVKGSEVPRSYKCTHTPTCPVKKKVERSPGHI 367
QY 279 TEIVYKGNHHPKQNTK-----
Db 368 TEIYKGAHNHPKPPNRRSALGSTNSLGLDQVDAEQGASGVNGDLQANFHKAPGGG 427
QY 297 ----RNSNSSLAIPHNSIRTEIPDQSYATHSGQWDSA--ATPENSISIGDDPEQS 351
Db 428 GFDWRNNRNDANLGEHCNR-----SAPFSAQNTRLESADVVSNSFNSENDEDEDRG 481
QY 352 SQKCKSGGDEDEPDAKRWKIEGENEGMSAPGSRVTRPVVVVQTTSDIILDGVRW 411
Db 482 THGVSQYEGEGESESERKLETYSADMTG--ATRAIREPVVVVQTTSEVTDIILDGVRW 540

```

```

QY 412 RYKQKVVKGNPNRPSRYKCTHTPGCPVRKHVERASHDLRAVITTYEGKHNHVPAAAGSG 471
Db 541 RYKQKVVKGNPNRPSRYKCTHTPGCPVRKHVERASHDLKSVITTYEGKHNHVPAAARNS - 599
QY 472 SHSVNRPMPNNAASNTHTAATSVRLLPVIHQSDNS--LQNR-----SQAPP 516
Db 600 SH-----VNSGASNTHPASVTAPAQNHILHRPEAAQLQANVARFDRQPSLGSFGLSGRPP 653
QY 517 E-QQSP-FTLEMLQSPG--SFGSGF-----GN-PMQSYVNOQQLSDNVFSSRRTKEEP 564
Db 654 QLGPNFGSYGMNQGGSLSLAMAGFHPNQNKPEVPMHPYLQGPRLMDHMGFMFPKEEP 713
QY 565 RRD-----MFLESL 573
Db 714 KVEPLSDPGPTAYQQFLNSL 733

RESULT 15
QY Q9ARD0 PRELIMINARY; PRT; 739 AA.
ID Q9ARD0
AC Q9ARD0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 79.8 kDa protein.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21178822; PubMed=11283350;
RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,
RA Schumacher K., Schmitt G., Schmidt R.;
RT "Comparative sequence analysis reveals extensive microcolinearity in
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella
RT genomes.";
RL Plant Cell 13:979-988(2001).
DR EMBL; AJ303343; CAC36397.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW Hypothetical protein.
SQ SEQUENCE 739 AA; 79809 MW; 6D97C8E64C1F221B CRC64;

Query Match 30.2%; Score 922; DB 10; Length 739;
Best Local Similarity 36.0%; Pred. No. 1.5e-56;
Matches 243; Conservative 96; Mismatches 182; Indels 154; Gaps 25;

QY 36 LASPLDNNKPPQGG-----LSERTGSGVPRF--KSTPPPSLPLSPPISSSYFAIPPLG 88
Db 78 LSPSPDQKMSRGGLLRIARAGFNAPKLNTESLRPAQLRQNGVRSP--YLTIPPLG 135
QY 89 SPALLDPSVLLNSNLPSTTCAFAVQS--FNWSSSGGNQIIVKEEDKSF-----SNF 142
Db 136 SPTLLSPVFLSNLVQSPSTTGKFLFSSGIDNRNSALMMDTDKRDNALESINSSSF 195
QY 143 SFQ-----TRS-----GPPASSTATYQSSNVTVQTO-----Q 169
Db 196 SFKVPETAPSLPFGTTTSSRLQVNPNSFQGGFPNIEVSHVSNLSLVRHEATQNPQT 255
QY 170 PWSFOEATKQDNFSSGKGMKNTENSSMQSFSPFIASVQTNHNGFQSDYGNYPQSQTL 229
Db 256 NGTLQSSDPRFSAEKDVM--ANNVLDSTFTQVGSADVHSPPLDE-----PQEDI 307
QY 230 SRR-----SDDGYNRWKYQKQVKGSENPRSYKCTYPCNCTPKKKVERSLDQI 278
Db 308 DQGGGDPNVAGAPAEADGYNRWKYQKQVKGSEVPRSYKCTHTPTCPVKKKVERSPGHI 367
QY 279 TEIVYKGNHHPKQNTK-----AIPHNSIRTEIPDQSY--ATHGSG 328
Db 368 TEIYKGAHNHPKPPNRRSALGSTNSLGLDQVDAEQGASGVNGDLQANFHKAPGGG 427

```

```
QY 329 QMD-----SAATPENSSISIGDDDFEQSSQKCKS 357
Db 428 GFDWRNNLDANLGSEHCNRSAPFSAQNNTLRLESGDAVDVSSNFSNDEDEDDRGTHGSVS 487
QY 358 GGDEYDEDEDAKRWKIEGENEGNSAPGSRVTVPRVVQTTSDIDILDDGYRWRKYGOK 417
Db 488 QGYEGEGDESESKRRKLETYSADMTG--ATRAIREPRVVQTTSEVDILDDGYRWRKYGOK 546
QY 418 VVKGNPNPRGYKCTHPGCPVRKHVERASHDLRAVITTYEGKHNHDVPAARGSGSHSVNR 477
Db 547 VVKGNPNPRGYKCTSHAGCNVRKHVERASHDLKSATITTYEGKHNHDVPAARNS-SH---- 601
QY 478 PMPNNASNHTNTAATSVRLLPVIHQSDNS-LQNR-----SQAPPE-QQSP 521
Db 602 --VNSGASNTHPASVTAPAQNHLPRAEAQLOQAMARFDRQPSLGSFGLSGRPPQLGPNP 659
QY 522 -FTLEMLQSPG--SFGSGF-----GN-PMQSYVNOQOLSDNVFSSRTKEEPRDD--- 567
Db 660 GFSYGMNQOQGLSLAMAGFHPNQNKPEVPMHPYLGQPRPMHDMGMFMFPKEEPKVEPLS 719
QY 568 -----MELESLL 574
Db 720 DPGFTAYQQFNSLL 734
```

Search completed: January 20, 2004, 16:17:51
Job time : 44 secs

THIS PAGE BLANK (USPTO)